(54) Title: COINCIDENCE CLONING METHOD AND LIBRARY

(57) Abstract

A method of obtaining common-sequence DNA fragments from two fragment mixtures, such as the fragments obtained from two different genomes, or from two different fragment-separation regions on a gel. The fragments of at least one of the two mixtures are modified such that heteroduplex fragments containing one strand derived from the first-mixture fragments and an opposite strand derived from homologous fragments in the second mixture can be isolated from homoduplexes formed by strand hybridization within each fragment mixture. The method can be used in applications relating to gene mapping, gene isolation, chromosome construction, cloning of conserved genes, and removal of repeat sequences from genomic DNA. Also disclosed are coincidence-sequence libraries formed by the method.

-1-

COINCIDENCE CLONING METHOD AND LIBRARY

1. <u>Field of the Invention</u>
        The present invention relates to methods for
obtaining DNA sequences which are common to two DNA frag-
ment mixtures derived from different sources, and to uses
of the method for gene mapping and cloning.

2. <u>References</u>

Beaucage, S.L., et al, <u>Tet Letters</u> (1981) 22:1859.

Britten, R.J., and Davidson, E.H., in <u>Nucleic Acid Hy-</u>
    <u>bridization. A Practical Approach</u>, B.D. Hames and
    S.J. Higgins, eds., IRL Press, Oxford, (1985), 3.

Britten, R.J., and Kohne, D.E., <u>Science</u> (1968), <u>161</u>:529.

Carle, G.F., et al, <u>Nuc Acids Res</u> (1984) <u>12</u>:5647.

Casna, N.J., et al, <u>Nuc Acids Res</u> (1986) 14:7285).

Duckworth, D., et al, <u>Nuc Acids Res</u> (1981) 9:1691.

Feinberg, A.P., et al, <u>Anal Biochem</u> (1983) 132:6.

Hames, B.D., and Higgins, S.J., eds., <u>Nucleic Acid Hy-</u>
    <u>bridization: A Practical Approach</u>, IRL Press, Oxford,
    (1985).

Kohne, D.E., et al, <u>Biochemistry</u> (1977) 16(24):5329.

Maniatis, T., et al, <u>Molecular Cloning: A Laboratory</u>
    <u>Manual</u>, Cold Spring Harbor Laboratory (1982), 280.

Matteucci, M.D., et al, <u>J Am Chem Soc</u> (1981) 103:3185.

Meselson, M., Stahl, F.W., <u>Proc Natl Acad Sci USA</u> (1958)
    44:671.

-2-

Schwartz, D.C., et al, <u>Cell</u> (1984), <u>37</u>:67.

Wasmuth, J.J., et al, <u>Am J Human Genet</u> (1986) 39:397.

3.   <u>Background of the Invention</u>

Recent developments in genetic mapping and clon-
ing have created a need for additional methods for
identifying and isolating genetic sequences from
chromosomes and chromosome regions of interest. Such
genetic sequences may then be used, for example, in
identifying the genes responsible for genetic diseases.

Methods provided in the prior art allow for the
isolation of genetic sequences from chromosomes and
chromosome regions of interest in certain circumstances.
Sorted chromosomes, isolated by physical methods from
various cell types, and cloned sequence libraries prepared
from sorted chromosomes, many of which are commercially
available (American Type Culture Collection, Rockville,
MD) contain genetic material from a selected chromosome,
and are available for most, although not all, human
chromosomes. While such sorted chromosomes have been
valuable in providing genetic sequences for regions of
interest in many cases, they do have some important
limitations. One is a relatively high level of contamina-
tion with nonspecific genetic material, which decreases
the utility of sorted chromosome material in isolating
sequences of interest. Another is that because the basis
of selection is at the level of the whole chromosome, it
is difficult to focus down on specific regions of inter-
est. This is in general true for tranlocations as well as
wild type chromosomes in the absence of any method for
specifically identifying and isolating the coincident
sequences between two sources of genetic material.

Subtractive hybridization techniques have proven
to be very valuable in isolating target genetic sequences
present in only one of two sources. This is useful, for
example, in isolating mRNAs (or the corresponding cDNAs)

-3-

which are expressed in various cell types after activation
or other stimuli. These methods rely on the use of two
cell sources which are largely identical, only one of
which contains the sequence of interest. Most often these
are mRNAs or the corresponding cDNAs, although genomic DNA
5   may also be used. Furthermore, these methods rely on the
use of an excess of sequences from the source which does
not contain the sequence of interest, in order to drive
the hybridization reaction towards the formation of
heteroduplexes.

10
4.   Summary of the Invention
          It is one general object of the invention to
provide a method which allows the cloning and identifica-
tion of DNA sequences which are common between different
15   sources of DNA fragments.
          It is another general object of the invention to
provide such a method which substantially overcomes
problems and limitations associated with the prior art, as
discussed above.
20   Another object of the invention is to provide a
variety of techniques which can be used to obtain such
common sequences, according to the method of the inven-
tion.
          Still another object of the invention is to use
25   such method to advance or solve problems in several areas
related to gene mapping, gene isolation, chromosome
construction, and identification of large groups of
conserved genes in humans and other mammalian species.
          Providing a novel method for removing repeated
30   sequences in a mixture of genomic fragments is yet another
object of the invention.
          The method of the invention is designed for
obtaining from a first mixture of DNA duplex fragments
derived from one source, those fragments which are
35   homologous to and end hybridizable with the duplex DNA

-4-

fragments in a second mixture of DNA fragments derived
from another source. According to the method, the frag-
ments are generated in a manner which allows heteroduplex,
end-hybridized fragments formed by the hybridization of
homologous DNA strands from the two DNA fragment mixtures

5 to be isolated from homoduplex fragments produced by
hybridization between opposite strands of the fragments in
the first or second mixture only, and from heteroduplex
fragments which are not end-hybridized.
Denatured strands from the fragments of the first and

10 second mixtures are reacted in a reaction mix under
hybridization conditions which yield heteroduplex, as well
a homoduplex, reaction fragments, and the end-hybridized
heteroduplex fragments are isolated from other nucleic
acid species contained in the reaction mix. The

15 hybridization reaction may be carried out with a molar
ratio of the two fragment mixtore or with a molar excess
of one of the mixtures.

In one general embodiment, the fragments are
generated in such a way that when the paired strands form-

20 ing the homoduplex fragments are mixed, denatured and
reannealed, end-hybridized heteroduplex fragments can be
isolated from other hybridization products by cloning,
based on a unique pair of ligatable ends in the desired
fragments. This method produces a library of cloned, co-

25 incident sequences which are enriched for single-copy
sequences, since the heteroduplex fragments with non-
hybridized ends are formed largely from repeated
sequences. The ligatable fragment ends in the end-
hybridized heteroduplex fragments may be generated either

30 when the fragments are produced, by restriction
endonuclease digestion, or by attachment of different
linkers to the two sets of DNA fragments. The linkers may
additionally contain methylated sites which allow
generation of unique end pairs in heteroduplex fragments,

35 by cutting with corresponding restriction endonucleases.

-5-

In another embodiment, one of the fragment
mixtures is modified with a label which allows physical
separation of heteroduplex fragments from homoduplexes.
The label may be an affinity label, such as biotin, which
allows separation of heteroduplex species based on (a)
5   initial binding to an affinity column and (b) subsequent
release of the unlabeled strand of the heteroduplex by
duplex denaturation. Alternatively, the label may be a
density label which permits physical separation of
heteroduplex from homoduplex strands based on density
10   gradient centrifugation.

In still another embodiment, the fragments in
the two mixtures are cloned in a vector which allows
expression of one fragment strand or its transcript from
one mixture, and the opposite fragment strand or its
15   transcript from the other mixture. Separation of the
heteroduplexes in this procedure is based on duplex forma-
tion and separation, for example, on a hydroxylapatite
column.

Considering the various applications of the
20   invention, one method of use is for cloning and/or analyz-
ing the gene sequences, and preferably the single-copy
sequences, which are carried on defined chromosomes or
chromosome regions. Here, the sources of the two DNA
fragment mixtures may be a two-species cell hybrid
25   containing the specified chromosome from one species, such
as a human/hamster hybrid containing a specified human
chromosome, and a cell line from the one species, such as
a human cell line. The procedure yields cloned, prefer-
ably single-copy, sequences present only in the defined
30   chromosome.

Another important application of the method is
for obtaining clones derived from a DNA fragment contained
in a mixture of fragments, such as are typically obtained
when DNA fragments are subfractionated, as by gel
35   electrophoresis. As an example, partial digest fragments

-6-

of genomic DNA, when fractionated by pulse field gel
electrophoresis, will yield several band regions contain-
ing a gene region of interest, as evidenced by the binding
of a selected probe to each of the regions of interest.
After eluting the fragments from each of two such gel
5 regions, these are then hybridized, according to the
method of the invention, to produce common-sequence
heteroduplex fragments derived from the desired probe-
binding fragments on the gel.

The method may also be used for identifying and
10 cloning gene sequences which are homologous between two
different species, e.g., humans and another primate spe-
cies. Because such homologous genes would be highly
conserved, they are likely to represent gene functions
which are important to the organism, such as gene func-
15 tions related to immunological defenses, peptide hormones,
and the like.

Another application of the method is for
identifying and cloning specific chromosomal regions, such
as the telomere regions at the end of chromosomes which
20 appear to be required for chromosome stability. The
method here involves cloning the coincident gene sequences
from hybrid cells, each of which contains the chromosomal
region of interest.

The method can also be used to enrich a mixture
25 of genomic DNA fragments for single-copy sequences, either
applied to a single DNA fragment mixture, such as total
genomic fragments from a given source, or in conjunction
with other applications mentioned above, in which the co-
incident fragments isolated by the method are enriched for
30 single-copy sequences.

In another aspect, the invention includes a
library of cloned DNA sequences produced by treating two
DNA fragment mixtures according to the method of the
invention, where the end-hybridized heterologous fragments
35 are cloned into a suitable cloning vector.

-7-

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read in conjunction with the accompanying drawings.

Brief Description of the Drawings

5

Figure 1 is illustrates the method of the invention, wherein heterologous duplex fragments are isolated from homologous fragments on the basis of different fragment ends present in the heteroduplexes;

10

Figure 2 illustrates the method in another embodiment wherein fragments in the two fragments are equipped with different linkers, and heteroduplex fragments are selected on the basis of different restriction sites formed by the two linkers at the opposite fragment

15

ends;

Figure 3 illustrates the method in another embodiment, in which the original homoduplex fragments are methylated at one of two different restriction sites, and heteroduplex fragments are isolated on the basis of unique

20

opposite-end restriction sites after digestion with endonucleases corresponding to the two methylase sites;

Figure 4 shows the method in a related embodiment, in which linkers attached to each of the mixtures of fragments contain two common internal restric-

25

tion sites, one of which is methylated, and different end sites, and heteroduplexes are distinguished from homoduplexes on the basis of different end sites which result after digestion with endonucleases specific for the internal linker sites;

30

Figure 5 illustrates another general embodiment of the method, in which heteroduplex molecules are isolated on the basis of binding to an affinity column and release of one strand of the heteroduplex on denaturation, where the released single-strand is contained in a cloning

35

-8-

vector which can be readily converted to double-strand
form;

Figure 6 illustrates an embodiment which is
similar to that in Figure 5, but where the released
single-strand material is annealed to form duplex frag-
5   ments which can be cloned into a suitable cloning vector;

Figure 7 illustrates the method of the invention
as it can be practiced using density-gradient centrifuga-
tion to separate heteroduplex from homoduplex fragments;

Figure 8 illustrates another method for carrying
10 out the method in which heteroduplexes are separated from
non-hybridizing DNA species by hydroxylapatite;

Figure 9 shows the steps in the application of
the invention to isolating clones from single fragments
obtained from a gel band region; and

15 Figure 10 illustrates the application of the
method to isolating a human chromosome telomere.

Detailed Description of the Invention

20
I.   Definitions

As used herein, the terms listed below have the
following meaning:

(a) Homologous fragments: Two DNA duplex frag-
25 ments are homologous with one another if the opposite
strands in the two fragments are capable of forming stable
duplex fragments under conditions of denaturation and
reannealing.

(b) End-hybridizable: Two fragments are end-
30 hybridizable if the homologous and opposite strands in the
two fragments are capable of hybridizing with one another
at their corresponding end regions by Watson-Crick base
pairing.   Such opposite strands are also said to be end-
hybridizable.

35

-9-

(c) End-hybridized: A fragment is end-hybridized if it is formed from end-hybridizable frag-ments. Typically, the strands forming an end-hybrized fragment will be hybridized along their entire lengths.

(d) Ligatable ends: The ends of a duplex frag-ments are ligatable if the fragment can be selectively incorporated into a cloning vector having defined ligation ends, in the presence of suitable ligation enzymes in vitro or in vivo. Ligatable ends include sticky ends, i.e., ends with short orverhang sequences capable of hybridizing with complementary overhang sequences, and blunt ends. Typically end-hybridized fragments will have ligatable ends.

(e) Homoduplex fragments: Homoduplex fragments are those formed by hybridization between homologous-fragment strands derived from the the same DNA fragment mixture.

(f) Hereroduplex fragments: Heteroduplex frag-ments are those formed by hybridization between homologous-fragment strands derived from different DNA fragment mixtures.

II. Preparing Coincidence-Clone Libraries

The method of the invention is aimed at obtain-ing gene sequences which are coincident in, i.e., common to the DNA fragments in two different mixtures of gene fragments. More particularly, the method is designed to obtain from the first mixture of duplex fragments, those fragments which are homologous to and end-hybridizable with the duplex fragments in the second fragment mixture.

These different mixtures can be obtained from different-species cells, from hybrid and non-hybrid cells, from different chromosomes or chromosome regions, from non-genomic sources, such as mitochondrial DNA, and in one embodiment, from the same DNA source, where the method is

-10-

used to obtain predominantly single-copy gene sequences
from the source. Various sources and methods of DNA
isolation are detailed in Part A.

5   In preparing the DNA fragment mixtures, at least
one of the mixtures is prepared in a manner such that when
a strand from one mixture is hybridized with a homologous,
end-hybridizable strand from the second mixture, the
resulting end-hybridized heteroduplex fragment has proper-
ties which allow its separation from homoduplex fragments
formed by hybridization between opposite strands of the
10 fragments in the first or second mixtures only, and from
duplex heteroduplex fragments which are not end-
hypridized. Part D below describes embodiments of the
method in which end-hybridized heteroduplex separation is
based on unique fragment ends which allow cloning into a
15 vector with selected insertion sites. In the embodiments
covered in Part C, the methods of separating heteroduplex
from homoduplex fragments involve physical separation of
labeled fragments. Part C describes another general ap-
proach to the invention, in which heteroduplex separation
20 is based on duplex formation from cloned, single-strand
species.

A.   Preparing DNA Fragment Mixtures

     The mixture(s) of duplex DNA fragments used in
25 the invention can be derived from a variety of multi-gene
DNA source(s), such as the genomic DNA from eukaryotic
cells or tissue samples, isolated chromosomes,
mitochondrial DNA, and subfractions of DNA obtained by
various DNA fragment separation procedures, such as gel
30 electrophoresis or centrifugation methods. The actual
source material used for DNA isolation may be whole cells,
or subfractionations thereof, such as cell nuclei, or
isolated chromosomes from cells.

     In many applications, the cell line used as the
35 DNA source for at least one of the fragment mixtures is a

-11-

hybrid cell line containing at least one chromosome or
chromosome region from one species, and the balance of the
chromosome material from one or more other species. These
hybrids may be obtained from known sources, or produced
according to published methods. For example, Example 1
utilizes as one source of DNA material, the genomic DNA
obtained from the somatic cell hybrid HHW661, a hamster-
human hybrid containing a translocation of human chromo-
some region 4p onto hamster chromosome 5 (Wasmuth). In
another application described in Section III below, the
two sources of DNA are both hybrid cells, one containing a
human chromosome 8, and another a human chromosome 4 with
a translocated portion of human chromosome 8.

Where the source of DNA is a biological sample,
the DNA can be isolated by standard procedures, which
typically include successive phenol and phenol/chloroform
extractions (Maniatis, p. 280). To illustrate, Example 1
describes the isolation of genomic DNA from two cell
lines. Where the DNA mixtures are derived from
subfractionated DNA fragments, such as from the agarose
gels, conventional methods of DNA extraction, such as
electroelution, gel maceration, or the like are used. The
elution of DNA fragments from agarose gel regions is
described in Example 9.

Typically, the isolated DNA is obtained in
relatively intact form, and is fragmented by digestion
with one or more selected restriction endonucleases, to
form the desired mixture of DNA fragments. In the usual
case, the DNA fragments in the mixture are formed by
complete digestion with one or more endonucleases, to
final fragment sizes of preferably between about 200 to
10,000 basepairs. Since in most applications, the
heteroduplex fragments formed in the method are cloned,
the upper size limit of fragments in the two mixtures is
limited to clonable fragment sizes, generally less than 40
kilobases, and preferably no more than 10-20 kilobases.

-12-

The choice of restriction endonucleases used in
forming the DNA fragments in the two mixtures will depend
on the specific approach used for isolating heteroduplex
fragments, as will be clear from the various approaches
described in Parts B-D below. In particular, the approach
5   used will dictate optimal fragment size and nature of the
cut ends. After endonuclease digestion to form the DNA
fragments, the fragments may be further modified by fill-
ing recessed ends, ligation of end linkers and/or
restriction-site methylation (Part B), by nucleotide
10 labeling (Part C), and/or by cloning into a single-strand
vector (Part D). Methods for performing such modifica-
tions are detailed in Examples 1-8 below.

B.   Heteroduplex Selection By Cloning
15   This general embodiment exploits differences in
end regions of the fragment hybridization products, to
selectively clone end-hybridized heteroduplex fragments
into a suitable cloning vector. In particular, the method
is effective to isolate heteroduplexes consisting of end-
20 hybridizable, homologous strands from homoduplex fragments
and from duplex fragments which are not end-hybridized,
i.e., which have one or more extended, non-hybridized end
regions. Since the preponderance of duplex fragments
which are not end-hybridized are formed by hybridization
25 between repeat sequences, the method is therefore effec-
tive in enriching for single-copy sequences which are co-
incident to the two fragment mixtures.
        One simple method of the invention which
exploits end-hybridized heteroduplex selection by cloning
30 is illustrated in Figure 1. Here two DNA fragment
mixtures, designated DNA-I and DNA-II, are each prepared
by digesting the corresponding DNA material to completion
with a restriction endonuclease, such as MboI, which
produces sticky fragment ends. One of these fragment
35 mixtures, e.g., the DNA-I mixture, is further treated with

-13-

Klenow fragment in the presence of the required
nucleotides, to fill in both recessed ends of the frag-
ments, forming blunt end fragments, as indicated.

The two fragment mixtures are now denatured and
reacted under hybridization conditions which yield
homoduplex and heteroduplex fragments. The hybridization
reaction may be carried out by traditional hybridization
methods, involving slow hybrid formation in a single-phase
aqueous or aqueous/formamide medium, at a reaction
temperature slightly below the melting temperature of the
duplex material, according to published methods (Britten,
1968; Britten, 1985; Hames). Preferably, however, the
hybridization reaction is performed according to a more
recent phenol emulsion reaction technique (PERT), or
formamide-phenol emulsion reaction technique (F-PERT),
which greatly accelerates the hybridization reaction
(Kohne, Casna). One potential drawback of the PERT
hybridization approach in this method is the potential for
DNA shearing during emulsification, resulting in blunt or
sticky ends which are unrelated to the original fragment
ends. This problem may be minimized by a light digestion
with S1 nuclease after ligating the hybridization products
into the cloning vector (below).

As shown in Figure 1, the hybridization reaction
produces three general classes of duplex fragments. The
first of these include original homoduplex fragments
formed by hybridization between end-hybridizable
homologous strands of the fragments in the first or second
mixtures only. These homologous duplex fragments have
either opposite blunt ends or opposite sticky ends. The
second class of fragments are homoduplex or heteroduplex
fragments formed from opposite strands which are not end-
hybridizable. Typically such fragments are formed from
imperfect copies of themselves, as is expected of repeat
sequences contained in a variety of different-size digest
fragments. At least one of the ends of these non-

-14-

hybridized fragments is irregular in that it has a
relatively long end-region of non-hybridized single-strand
DNA. The third class of fragments are end-hybridized
heteroduplex fragments. As seen in the mid portion of
Figure 1, these fragments have opposite sticky and blunt
5 ends.

The fragment mixture formed by reacting the op-
posite strands from the first and second DNA mixture are
now cloned into a cloning vector which is designed to in-
corporate selectively only those duplex fragments having
10 opposite sticky and blunt ends, i.e., heteroduplexes
formed from end-hybridizable homologous strands.
Typically, the vector is a plasmid, such as the pUC18
plasmid illustrated in Figure 1, which is cut at a
polylinker site to expose ends which are compatible with
15 the sticky and blunt ends.of the desired heteroduplex
fragments. The reaction fragments are ligated into the
vector after removal of the small polylinker segment.
Selection of successful recombinants, on a suitable host,
is carried out by conventional methods. Since some of the
20 end-hybridized heteroduplex fragments may be formed from
end-hybridizable homologous strands, the successful re-
combinants may be further screened with labeled repeat
sequence to eliminate the small percentage of repeat
sequences.

25   This method is detailed in Example 1, which
generally follows the reaction scheme shown in Figure 1.
Three out of five clones which were screened were
heterologous fragment inserts, i.e., derived from
sequences common to both genomic DNA sources. Only one of
30 the 48 clones which were screened by repeat-sequence
probes showed evidence of repeat-sequence fragments.

In a related method, the two DNA fragment
mixtures are prepared by (a) digesting the first and
second DNA with different endonucleases, such that the
35 first and second fragment mixtures have different sticky

ends. The nucleases used are selected such that the
hybrid sticky ends formed by hybridization between first-
and second-mixture equal-size strands are different from
either of the homoduplex sticky ends in the original frag-
ment mixtures.

Figure 2 illustrates another approach for
generating fragment mixtures in which homoduplex and end-
hybridized heteroduplex fragments can be separated by an
appropriate cloning vector. Here the characteristic
sticky ends used to distinguish homoduplex from end-
hybridized heteroduplex fragments are provided by end
linkers which are attached to the original digest frag-
ments. With reference to Figure 2, the DNA material from
the two sources is originally digested to completion with
an endonuclease, such as MboI, which preferably cuts the
DNA at relatively frequent intervals, e.g., every 200-
5,000 basepairs. The first fragment. mixture is then mixed
with one linker, designated linker I in the figure, which
is designed for attachment to the fragment sticky end and
provides an internal, preferably infrequent restriction
site, such as the XhoI site indicated. Treatment of this
fragment mixture with the linker-site endonuclease now
yields relatively small fragments with opposite rare-
cutter site ends. A second linker, similarly designed for
attachment to the original digest mixture and carrying a
second internal and preferably infrequent endonuclease
site, such as NotI (linker II in Figure 2) is similarly
attached to the second fragment mixture, which is then
treated with the linker-site endonuclease, to generate a
second fragment mixture composed of small fragments with
infrequent-site sticky ends. In addition to the require-
ment of a selected restriction-site sequence, the linker
sequences are also designed for hybridization with one
another, as illustrated in linkers I and II in Figure 2.

The two fragment mixtures are now mixed, de-
natured, and reannealed, as above, to produce hybridized

-16-

fragments consisting of end-hybridized and non-end-
hybridized homoduplex and heteroduplex fragments. The
end-hybridized homoduplex fragments have opposite sticky
ends which are either both linker I or both linker II
ends; non-end-hybridized homoduplex and heteroduplex frag-
ments have at least one irregular end; and end-hybridized
heteroduplex fragments have one linker I end and an op-
posite linker II end, as indicated. These fragments are
mixed into a cloning vector which selectively incorporates
the linkerI/linker II ends, and successful recombinants
are selected as above.

In the scheme illustrated in Figure 2, and
described in Example 2, the two fragment mixtures are
prepared with XhoI and NotI sticky ends, and the hybrid-
ized fragments are cloned into the XhoI/NotI site of a
Bluescript® plasmid.

Figure 3 illustrates another procedure for
preparing the fragment mixtures for selection of
heteroduplex fragments on the basis of hybridized-end
characteristics. This procedure utilizes methylation at
internal restriction sites, followed by endonuclease
treatment of the hybridization products, to generate
unique fragment ends in equal-size heteroduplexes.

The DNA fragment mixtures are initially prepared
by complete digestion with a one or more selected
endonucleases, where the endonuclease(s) used is selected
to produce preferred fragment sizes of at least about
1,000-2,000 kilobases, to insure that most of the
fragmenst contain internal frequent-cutting sequences,
such as AluI and HaeIII sequences. For illustrative
purposes, the fragments shown in Figure 3, which are
produced by BamHI (B) digestion, contain a single internal
AluI (A) and two HaeIII (H) restriction sites. The first
fragment mixture, designated DNA-I in the figure, is
treated with a selected methylase, such as AluI methylase,
to methylate both fragment strands at one frequent-cutting

-17-

site, as indicated by the "*" symbols in the figure.
Similarly, the second fragment mixture is treated with a
second methylase, such as <u>Hae</u>III methylase, to methylate
both fragment strands at a second frequent-cutting site in
the strands.

5   The two fragment mixtures are mixed, denatured,
and reannealed, as above, to produce hybridized fragments
consisting of both homoduplex and heteroduplex fragments.
With reference to Figure 3, it is seen that the homoduplex
fragments (whether or not formed from end-hybridizable
10   strands) are methylated at both strands at one frequent-
cutting site only, and thus can be digested by
endonuclease cutting at the non-methylated frequent-
cutting site. By contrast, the heteroduplex fragments
(again, whether or not formed from end-hybridizable
15   strands) are methylated on one strand or the other at both
frequent-cutting sites, and therefore protected against
endonuclease digestion by endonucleases which require
either frequent-cutting sequence.

       Digestion of the homoduplex fragments with
20   endonucleases, such as <u>Alu</u>I and <u>Hae</u>III, which cut at the
two frequent-cutting sites, will cleave all homoduplex
fragments (at the non-methylated frequent-cutting sites,)
but leave the heteroduplex fragments intact. As a result,
only those duplex fragments which (a) are formed of end-
25   hybridizable strands, and are either (b) heteroduplex
fragments or (c) homoduplex fragments which contain
neither of the above frequent-cutting sites, will retain
the original sticky ends, e.g., the two opposite <u>Bam</u>HI
ends, used in generating the two fragment mixtures.

30   Following digestion with the two frequent-
cutting endonucleases, the fragments are cloned into a
vector which selectively incorporates fragments with
original sticky ends. Successful recombinants are
selected as above. Clones which do not contain either of
35   the two frequent-cutting sequences used above, and which

are therefore suspect as being derived from homoduplex fragments, can be identified by resistance of the plasmids to linearization by digestion with either of the two endonucleases. Details of the method are given in Example 3.

A fourth method of heteroduplex selection by cloning employs elements of both the end-linker and site-methylation approaches just described. In this method, which is illustrated in Figure 4, fragment digestion and attachment of different linkers (linkers I and II in the figure) are carried out substantially as in the method illustrated in Figure 2. Here, however, the linkers contain, in addition to the "proximal" sticky end used for ligation to the fragments, such as an MboI sticky end, and a rare-cutting sequence near the "distal" linker end, such as a NotI sequence, two "internal" restriction sequences, in the present example, AluI, and HaeIII sites. The two internal-site sequences are referred to more generally as A and B sequences, and the distal-site sequences, such as NotI and XhoI sequences, as C and D sequences. Thus linker I in the figure has the sequences A/B/C and linker II, the sequences A/B/D.

Following linker attachment, the DNA-I fragment mixture, having the linker-I ends, is treated with a methylase which is specific for the A linker sequence, and the DNA-II fragment mixture, having the linker-II ends, is treated with a methylase specific for the B linker sequence. The resulting fragment mixtures are methylated at both linker strands, at either the A or B sequence and at any A or B internal sequences in the fragments, as indicated in the figure.

The two fragment mixtures prepared as above are now mixed, denatured and annealed, as above, to produce (a) end-hybridized homoduplex fragments which are protected at one or the other but not both of the A or B linker sequences, (b) non-end-hybridized homoduplex and

heteroduplex fragments having at least one irregular end and (c) end-hybridized heteroduplex fragments which are protected at both A and B linker sequences, by virtue of different-strand methylation in the linker region, and having opposite-end C and D sequences. Digestion of the

5   reaction fragments with endonucleases specific for both A and B sequences cuts the homoduplexes at all A or B sequences, producing fragments with either A-sequence or B-sequence opposite sticky ends. Heteroduplexes, by contrast, are not cut by either endonuclease, and thus

10   retain their opposite C and D sequences. Further diges- tion with endonucleases specific for C and D sequences now produce C and D sticky ends in the opposite ends of end- hybridized heteroduplex fragments. It can be appreciated that a small percentage of fragments containing internal C

15   or D sequences may have opposite C or opposite D sticky ends.

The digest fragments are now cloned into a suit- able vector containing C and D sticky end sites, and the successful recombinants selected as above. The fragments

20   may also be cloned into vectors containing opposed C- sequence sticky ends, or D-sequence sticky ends, to clone those heteroduplex fragments containing internal C or D sequences. Example 4 details a procedure which follows the general scheme shown in Figure 4.

25   As can be appreciated from the above, all of the procedures presented above share a number of common features and advantages:

(a) In all procedures, the two fragment mixtures are generated from the associated DNA source in such a way

30   that the hybridization products produced by reacting the two fragment mixtures under hybridization conditions can be separated on the basis of selective incorporation into a suitable cloning vector. Thus the method for isolating the desired heteroduplex fragments also yields a fragment

35

-20-

library which is enriched for end-hybridizable, coincident sequences.

(b) All of the procedures involve relatively simple digestion and methylation and/or linker attachment manipulations in generating the fragments.

5   (c) All of the procedures effectively select against repeat sequences, by virtue of the irregular fragment ends which are generally associated with repeat sequences.

C. Heteroduplex Selection Based on Physical Properties

10   In the various embodiments of the invention described in this section, heteroduplex fragments are separated from homoduplex fragments on the basis of a physical property related to a nucleotide label. The label may be either a density label, such as an ^{15}N-

15 labeled nucleotide, or an affinity label, such as biotin, which is incorporated into both strands of one fragment mixture. Heteroduplex fragment separation then involves isolating fragments containing one labeled and one unlabeled strand from completely labeled or completely

20 unlabeled homoduplex fragments.

An example of one approach using an affinity label for heteroduplex separation is illustrated in Figure 5. Here one fragment mixture, designated DNA-I, is labeled in both strands with biotin. The other mixture,

25 designated DNA-II, is cloned into a vector, such as M13, which can be grown in single strand form. Because the cloning vector is used as a source of one strand only (either the sense or anti-sense strand), the original fragments are prepared by digestion with two

30 endonucleases, such as EcoRI and HindIII, so that the fragments can be introduced directionally into the vector. The DNA-I fragments are prepared by digestion with the same pair of enzymes.

35

-21-

The method illustrated in Figure 5 does not involve a cloning step which discriminates against unequal-size heteroduplex fragments. Therefore, in order to ensure that the heteroduplexes are predominantly end-hybridized, one of the two mixtures, and preferably the

5 DNA-I fragment mixture, is initially treated to remove repeat sequences. This can be done by conventional slow hybridization techniques carried out in a single-phase reaction system, as referenced above. Typically, the denatured fragments in the mixture are hybridized to an

10 initial $C_o t$ value at which most of the repeated sequences are hybridized, and most of the single-copy sequences are still in single-strand form. After removing the hybridized species by binding to a hydroxyapatite column, the single-strand material is carried to a second $C_o t$ value at

15 which the single-copy strands are predominantly hybridized. This general techniques is detailed in Example 5A.

Alternatively, several methods described herein for selecting coincident species may also be applied initially to removing repeat sequences, as will be

20 considered in Section III below.

The single-copy fragments from above are now labeled with a biotin label, according to one of the general procedures detailed in Example 5B. All of the methods produce labeling of both fragment strands, as is

25 required. Although biotin is the preferred affinity label, any label which can be incorporated into polynucleotides and which has a binding partner capable of binding the label specifically and with high affinity may be used. The affinity label is also referred to herein as

30 an epitopic molecule, and the binding partner, as a binding molecule. Exemplary binding pairs of epitopic molecule/binding molecule include biotin/avidin, biotin/streptavidin, antigen/antibody, and carbohydrate/lectin. The methods described in Example 5B for incorporation of

35 biotinylated nucleotides into polynucleotide fragments are

-22-

generally applicable to incorporation of nucleotides
derivatized with other epitopic molecules.

The labeled, single-copy strands are now mixed
with the cloning vector containing the DNA-II fragment
inserts and grown under conditions which yield one vector
strand (sense or anti-sense) only. The mixture is de-
natured and allowed to reanneal, as above. With reference
to Figure 5, the annealing reaction produces homoduplex
fragments, heteroduplex fragments consisting of a labeled
fragment strand from the DNA-I mixture, and the homologous
DNA strand from the cloned DNA-II mixture, and single-
strand species from both mixtures (not shown). These re-
action products are now applied to an affinity support
material having surface-bound binding molecules, to bind
all labeled duplex fragments to the support, with elution
of non-hybridized DNA-II strands. The support-bound
material is now denatured, either by heating, raising pH,
and/or addition of denaturing solvents, such as a water/
formamide mixture, to release the non-labeled, cloned,
single strand material from the support. The resulting
phage material is used to transfect a suitable bacterial
host, and grown in either single-strand or double-strand
form. This method is detailed in Example 5.

A related method which does not require removing
repeat sequences from one of the fragment mixtures is il-
lustrated in Figure 6. Here both fragment mixtures are
generated by digestion with the same endonuclease, and one
of the fragment mixtures is labeled, as indicated. The
labeled and unlabeled mixtures are now mixed, denatured,
and reannealed, as above, producing homoduplex fragments
with both or neither fragments labeled, and heteroduplex
fragments with one strand only labeled.

The hybridization products are passed through an
avidin or streptavidin column, binding labeled homoduplex
and heteroduplex fragments to the column, with elution of
the unlabeled homoduplex fragments. The bound fragments

-23-

are now denatured, as above, and the unlabeled single-
strand species are eluted. It will be appreciated that
the eluted DNA strands are (a) all derived from the
unlabeled fragment mixture, (b) represent both end-
hybridizable and non-end-hybridizable strands, and (c)
5   include both sense and anti-sense strands. These single
strand species are ethanol precipitated, and reannealed,
forming homoduplex fragments which are derived from
heteroduplex fragments only, i.e., are all coincident with
fragments in the labeled fragment mixture.

10   With continued reference to the figure, the
reannealed end-hybridized duplex fragments (representing
predominantly single-copy fragments), contain the same
sticky ends as the original unlabeled fragments, whereas
the duplex fragments which are not end-hybridized contain
15   at least one irregular end. The total fragments are mixed
with a suitable cloning vector which selectively in-
corporates the regular sticky end fragments, with selec-
tion for successful recombinants as above. The method is
detailed in Example 6, where the fragment mixtures are
20   formed with MboI digestion, and the reannealed unlabeled
fragments are cloned into the MboI site of a pUC18 vector.

          Figure 7 illustrates a method of density gradi-
ent separation of heterologous and homologous fragments.
Here the two fragment mixtures are prepared by digestion
25   with a frequent-cutting endonuclease, such as MboI, and
one of the fragment mixtures is labeled, as above, by in-
corporation of a heavy isotopic nucleotide, such as ^{15}N-
labeled nucleotides, where the label may be carried in one
or more of the nucleotide species. Incorporation of the
30   labeled nucleotides is by one of the methods detailed in
Example 5B for incorporation of biotinylated nucleotides
into duplex DNA.

          The labeled and unlabeled fragments are mixed,
denatured and reannealed as above, yielding homoduplexes
35   with both or neither unlabeled strands and coincident

-24-

heteroduplex fragments with one labeled and one unlabeled
strand.   These three species of duplex fragments are then
fractionated by equilibrium density centrifugation, ac-
cording to classical procedures, such as on a CsCl gradi-
ent.   In the density gradient shown in Figure 7, where the
5   labeled strands are shown by wavy lines, the heteroduplex
fragments fractionate between the lighter unlabeled
homoduplexes and the heavier, fully labeled homoduplexes.
The heteroduplex fraction is recovered by aspiration.
This fraction contains both end-hybridized and non-end-
10   hybridized fragments, and the former are isolated by clon-
ing into an appropriate cloning site in a plasmid vector,
as in the method immediately above.   Details of this
method are given in Example 7.

15   D.   Coincidence Fragment Selection by Duplex Formation
          The method presented in this section relies on
the formation of duplex fragments from homologous sense
and anti-sense strands contributed by the first and second
DNA fragment mixtures, respectively.   With reference to
20   Figure 8, each of the two fragment mixtures is initially
prepared by digestion with two selected endonucleases,
such as EcoRI and HindIII, producing fragments which can
be inserted in an oriented fashion in a cloning vector
which can be grown in either a single-strand or double-
25   strand form.
          In a preferred approach, the two fragments
mixtures are cloned into a pair of cloning vectors which
are designed to receive fragments in one or two defined
orientations, in a double-strand form, and which therefore
30   produce opposite insert strands, in a single-strand form.
One such vector pair includes the vectors M13/mp18 and
M13/mp19 which have polylinkers arranged in opposite
orientations, for accepting inserts cut with a pair of
selected endonucleases, such as EcoRI and HindIII, in op-
35   posite orientations.   The cloning step is shown in Figure

-25-

8, where the first EcoRI/HindIII fragment mixture is cloned into an mp19 plasmid in one orientation, and the second EcoRI/HindIII fragment mixture is cloned into an mp18 plasmid in the opposite orientation. When these plasmids are grown under single-strand phage conditions, the mp19 vector produces the sense (+) strand of the

5 insert, and the mp18 vector, the anti-sense (-) strand.

The two single-strand phage mixtures, containing the opposite-strand inserts, are now mixed and annealed, preferably using the above F-PERT procedure, yielding phage complexes having opposite strand duplex regions. As

10 indicated in Figure 8, phage complexes formed from end-hybridizable inserts allow end region annealing of the opposite-strand polylinker sequences present in the two cloning vectors, so that the duplex inserts are bounded by defined duplex restriction sequences, and in particular,

15 the sequences used for inserting the original fragments into the double-strand vectors. In the example il-lustrated in Figure 8, and detailed in Example 8, these sequences are those recognized by EcoRI and HindIII. By contrast, opposite strand complexes formed from non-end-

20 hybridizable fragments have at least one irregular mismatch at the insert end which prevents annealing at the vector polylinker sequences.

The annealed fragments are now digested with endonucleases which cut at opposite vector polylinker

25 sites, and preferably at the sites used to introduce the original fragments into the two vectors, to avoid cutting the inserts themselves at internal sites. Thus, in Figure 8, the EcoRI and HindIII sites used to generate the original fragment mixtures, and to introduce the fragment

30 mixtures into the two cloning vectors, are also used to digest the duplex phage species. The resulting digest fragments are then cloned into a suitable cloning vector, such as pUC18 opened at its EcoRI and HindIII sites, which

35

selectively incorporates the equal-size duplex fragments.
This method is illustrated in Example 8.

This approach has the potential for greater
discrimination between coincident and non-coincident
sequences, since only coincident sequences form hybrid

5   duplexes, and therefore could be introduced into a duplex
cloning vector. The method also has the potential for
good discrimination between end-hybridizable and non-end-
hybridizable duplexes, since only equal-size duplexes are
released from the hybridized products in a clonable form.

10   The limitation of the method is the need for two cloning
steps, one in forming the single-strand fragment mixtures,
and the second in selecting single-copy annealed
hybridization products.

15   III.   Applications

This section discusses applications of the co-
incident cloning method to gene mapping, gene isolation,
chromosome construction, cloning of conserved genomic
sequences, and removing repeat sequences from genomic DNA.

20

A.   Cloning Single-Copy Chromosome-Region Sequences

There are a variety of applications in which it
is useful to identify and clone coincident single-copy
sequences from different DNA sources. For example, it

25   would be useful to clone all of the single-copy gene
sequences from a given human chromosome or chromosome
region. In many instances it is not possible to isolate
the chromosome or chromosome region of interest, either
because of limitations of physical isolation or because

30   the chromosome region of interest is not mapped.

As one example of this application, assume the
problem is to clone all of the single-copy sequences in
human chromosome 4 (C4), for purposes of constructing a
library of probes for C4. As a first step, one would

35   first construct a human/non-human hybrid containing C4 on

-27-

a background of non-human chromosomes. Typically the
hybrid would be a mouse/human or hamster/human hybrid
containing a single C4 chromosome. To clone the C4
single-copy sequences, restriction fragments from this
source (DNA-I) and from the entire human genome (DNA-II)
are mixed, and reacted under hybridization conditions,
according to one of the methods from Section II above, to
produce hybridization products representing coincident
sequences, i.e., sequences associated with C4. The
hybridization products are further treated and cloned, as
above, to yield a library of cloned, C4 sequences enriched
for single-copy species. These clones in turn can be
radiolabeled to provide a substantially complete bank of
probes for human C4. This application is illustrated
particularly in Example 1, which demonstrates the cloning
of single-copy species associated with human chromosome 5
containing a translocation of chromosome 4p.

It will be appreciated how the method can be
similarly applied, in the construction and analysis of
hybrid genomes, to answering questions about (a) how much
of the total genome in a hybrid cell is contributed by a
selected species, (b) which chromosome(s) or chromosome
segment(s) are present from the selected species, and (c)
changes in chromosome composition over time.

B.   Cloning Sequences of Single Genomic Fragments
An important problem in human genetic studies is
identifying genes or gene groups which are related to
particular genetic diseases. Often the search for such
genes begins by screening human single-copy probes, using
restriction fragment polymorphism analysis to identify
probes which are associated with a disease related
restriction pattern. These probes presumably correspond
to regions which are genetically linked to the disease-
related gene of interest, but which may still be up to
1,000 kilobases or more from the gene of interest. Once

-28-

these probes are identified, they may be used to probe
large genomic digest fragments which can be size-
fractionated by new gel electrophoresis methods, such as
pulsed-field gel electrophoresis (PFGE), which provide
greater resolution of large genomic fragments by virtue of
5   orthogonally disposed electrical fields (Smith).

In theory, PFGE can be used to fractionate large
genomic fractions, and the fragments of interest, i.e.,
those associated with one or more identified probes, can
be identified on the gel by probe binding techniques, such
10   as Southern blotting. The limitation of this approach is
the relatively large number of same-size digest fragments
which will typically be found in a probe-binding gel
region. That is, elution of the digest fragments from a
probe binding region may yield many distinct fragments,
15   without any practical way of resolving and isolating the
probe-binding fragment of interest. Accordingly, efforts
to map the fragment region of interest with cloned library
subfragments would be quite difficult, since most of the
cloned subfragments would not relate to the fragment of
20   interest.

The application of the present invention to this
problem is illustrated in Figure 9. Here the duplex DNA
shown at the top in the figure represents a segment of DNA
containing a probe-binding region P which is adjacent a
25   gene region of interest G where both P and G are located
between a pair of restriction sites S_3/S_4. The restric-
tion sites S_i are preferably at least about 100 kilobases
from one another. The objective is to clone fragments in
the S_3/S_4 fragment segment only, for purposes of further
30   mapping the relationship between P and G and identifying
one or more cloned G subfragments.

As a first step in the method, the DNA is
partially digested with the endonuclease which cuts at the
rare S sequences. Methods for forming partial DNA digests
35   which are suitable in the present method are given in

-29-

Example 9. As seen, the partial digest produces a number of different-size fragments which contain the desired S_3/S_4 segment, including the S_3/S_4 fragment. The partial digest fragments are now fractionated by PFGE, substantially according to methods described and

5 referenced in Example 9, and the gel is examined for probe-binding regions (containing the S_3/S_4 fragment) by Southern blotting, using the previously selected probe. Two of the probe binding regions are now removed and the digest fragments are eluted. For purposes of illustra-

10 tion, it is assumed that the probe-binding regions identified as S_1/S_4 and S_3/S_4 are so identified and eluted. In particular, it is an advantage to select as one of the probe-binding region, the region containing the smallest probe-binding region, presumably representing the smallest

15 digest fragment possible, in this case, the fragment S_3/S_4.

The two eluted gel fractions are used as the two DNA sources from which coincident sequences can be cloned, according to the method of the invention. Each of these

20 fragment mixtures is digested to completion with one or two selected endonucleases and prepared for hybridization, according to one of the methods detailed in Section II. Hybridization and cloning of heteroduplex fragments formed from end-hybridized strands yield cloned subfragments

25 which are common to both elutate mixtures. Assuming that the S_3/S_4 fragment contains the only sequences common to both eluates, the method yields clones containing only sequences present in the S_3/S_4 fragment, and enriches single-copy sequences. With this limited library, mapping

30 of and gene identification in the S_3/S_4 fragment is greatly simplified.

It will be appreciated that the method is similarly applicable with other recombinants for generating fragments that fractionate in different parts of the

35 gel, or in more than one gel, which contain coincident

-30-

sequences. This may be accomplished, for example, by
using a source or sources containing a restriction
fragment-linked polymorphism for the rare cutter enzyme S
in the region of interest, or by cutting with two differ-
ent rare cutter enzymes.

C. Cloning Conserved Sequences

    It is now recognized that many of the functional
genes in higher organisms have been relatively conserved
during evolution, as evidenced by considerable sequence
homology between analogous-function genes in related
organisms. In general, greater conservation is seen in
more closely related species, and also with more
important, i.e., fundamental gene products, such as
histones and hemoglobin. Because conserved gene sequences
are likely to represent the most important functional
genes in an organism, it would be advantageous to obtain
all of the conserved sequences of an organism,
particularly in humans, in cloned form.

    In practicing the method, two genomic sources
from related species are selected. For cloning human
conserved genes, a primate species such as lemur would be
preferred, since a more closely related species, such as
chimpanzee, may give too much general gene homology. The
two DNA sources are fragmented, denatured, reannealed and
cloned, according to one of the procedures in Section II,
yielding a library of conserved sequences enriched for
single-copy sequences.

D. Cloning Human Telomere Regions

    To date, efforts to identify telomere sequences
in human chromosomes have not been successful, despite the
importance of this region for chromosome stability.
Cloned telomere sequences may be important, for example,
for constructing stable chromosomes which can be used for
gene therapy.

-31-

One approach to cloning telomere sequences, according to the present invention, is outlined in Figure 10. The upper portion of the figure shows a known translocation in the end regions of human chromosomes 4 and 8 in which end portions of the two chromosomes,

5   including the telomere region, are exchanged. The objective of the method is to clone those sequences, presumably including telomeric sequences, which are present in the C8 translocation on the C4 chromosome.

As a first step in the method, hybrid cells

10   containing in one case chromosome 8, and in the other case, chromosome 4 with the chromosome 8 translocation are produced. As an example, one hybrid cell is a Chinese hamster ovary (CHO) cell containing the C4/C8 translocation chromosome, and the other hybrid is a mouse

15   cell containing a normal human C8 chromosome, as indicated in the figure. DNA from the two cell types is isolated, and fragmented as above, to form the two DNA fragment mixtures used in the method. The coincidence sequences, which include those single-copy sequences derived from the

20   translocated portion of C8, as well as those sequences conserved between Chinese hamsters and mice, are obtained by one of the coincidence methods discussed in Section II. Those clones containing rodent conserved sequences are then identified and removed by screening with total DNA

25   from either rodent cell.

E.   Cloning Infectious Microorganisms

This application is aimed at cloning DNA sequences derived from infectious microorganisms which (a)

30   have not yet been identified and isolated, and (b) are infectious toward disparate hosts, such as humans and rodents. The two infected cell types from the disparate hosts are used to produce the two DNA fragment mixtures from which coincident sequences will be derived, according to the method of the invention. The library of cloned

35

-32-

sequences may be further screened with the sequences
derived from the two host sequences, such as human and
hamster genomic sequences, to remove host sequences from
the library. The remaining cloned fragments now represent
sequences derived from the infectious agent. These
5   clones, in turn, can be used as probes for identifying the
infection in cells, or for determining sequences in the
genome in the infectious agent, for purposes of preparing
diagnostic or vaccine reagents.

10   F.   Enriching Genomic Fragments for Single-Copy Sequences
              As indicated above, the method of the invention
may also be used for removing repeat sequences from
genomic DNA, to enrich a genomic fragment mixture for
single-copy sequences. This application is based on the
15   ability of the method to discriminate against heterologous
fragments formed from non-end-hybridizable strands, as-
sociated predominantly with repeat sequence hybrids.
              In practicing the method, the genomic material
of interest is divided into two portions, and each of
20   these is then used in generating the two fragment mixtures
which are to be hybridized. The two mixtures are reacted
under hybridization conditions which yield heteroduplex
fragments, as discussed above, and these are further
cloned to selectively remove fragments formed from non-
25   end-hybridized fragments. The resulting genomic library
can be further screened with known repeat sequences to
further enrich the library for single-copy sequences.
              From the foregoing, it can be appreciated how
various objects of the invention are met. The method of
30   the invention provides a simple, practical approach for
selecting out of two large mixtures of genomic fragments,
those coincident sequences which are common to both
mixtures. In particular, the method typically yields a
library of cloned coincident sequences which are enriched
35   for single-copy sequences. The method may be performed by

-33-

a variety of procedures which rely on fragment end
characteristics, physical properties, and/or duplex forma-
tion in cloned single-strand form.

      The method can be applied usefully to a number
of significant problems in genetic mapping and gene clon-
ing, including the specific applications described in this
section.

      The following examples illustrate methods of
coincidence cloning using heteroduplex cloning and/or
physical selection methods according to the invention, and
applications of coincidence cloning to selection of
specific genomic sequences. The examples are intended to
illustrate, but not limit, the scope of the invention.

Materials and Methods

      M13/mp18 and M13/mp19 are obtained from New
England Biolabs (Beverly, MA). Cloning plasmid pUC18 and
its host E. coli strain JM103 are obtained from Pharmacia.
Bluescript® cloning vector containing NotI and XhoI clon-
ing site is supplied by Stratagene (San Diego, CA).

      Terminal transferase (calf thymus), alkaline
phosphatase (calf intestine), polynucleotide kinase,
Klenow reagent, and S1 nuclease are all obtained from
Boehringer Mannheim Biochemicals (Indianapolis, IN); SP6
and T7 polymerase, from Promega Biotech (Madison, WI); and
proteinase K, RNase and DNase, from Sigma (St. Louis, MO);

      NotI, XhoI, SmaI, BamHI, HindIII, EcoRI, T4 DNA
ligase and T4 DNA polymerase, SalI, HaeIII, AluI, NotI
methylase, XhoI methylase, HaeIII methylase, and AluI
methylase are obtained from New England Biolabs (Beverly,
MA); oligo dT primer and oligo dA and oligo dT cellulose,
from PL Biochemicals (Milwaukee, WI); Chelex-100, from
Bio-Rad (Richmond, CA); Sephadex G-50, from Pharmacia
(Piscataway, NJ); and streptavidin agarose, from Bethesda
Research Labs (Bethesda, MD). Low-gelling temperature

-34-

agarose is obtained from Sea Plaque, FMC, and proteinase K
and phenylmethylsulfonyl fluoride (PMSF), from standard
sources. Nitrocellulose filters are obtained from
Schleicher and Schuell.

5   Synthetic oligonucleotides for vector modifica-
tions to introduce NotI and SfiI linkers are prepared by
conventional phosphotriester methods (Duckworth) or the
phosphoramidite method as reported (Beaucage; Matteucci),
and can be prepared using commercially available automated
oligonucleotide synthesizers. Alternatively, custom
10   designed synthetic oligonucleotides may be purchased, for
example, from Synthetic Genetics (San Diego, CA).
Kinasing of single strands prior to annealing or for
labeling is achieved using an excess, e.g., approximately
10 units of polynucleotide kinase to 1 nmole substrate in
15   the presence of 50 mM Tris, pH 7.6, 10 mM $MgCl_2$, 5 mM
dithiothreitol, 1-2 mM ATP, 1.7 pmoles gamma-^{32}p-ATP (2.9
mCi/mmole), 0.1 mM spermidine, 0.1 mM EDTA.

Site specific DNA cleavage is performed by
treating with the suitable restriction enzyme (or enzymes)
20   under conditions which are generally understood in the
art, and the particulars of which are specified by the
manufacturer of these commercially available restriction
enzymes. See, e.g., New England Biolabs, Product Catalog.
In general, about 1 ug of plasmid or DNA sequence is
25   cleaved by one unit of enzyme in about 20 ul of buffer
solution; in the examples herein, typically, an excess of
restriction enzyme is used to insure complete digestion of
the DNA substrate.

Incubation times of about one hour to two hours
30   at about 37°C are workable, although variations can be
easily tolerated. After each incubation, protein is
removed by extraction with phenol/chloroform, and may be
followed by ether extraction, and the nucleic acid re-
covered from aqueous fractions by precipitation with
35   ethanol (70%). If desired, size separation of the cleaved

-35-

fragments may be performed by polyacrylamide gel or
agarose gel electrophoresis using standard techniques. A
general description of size separations is found in
Methods in Enzymology (1980) 65:499-560.

Restriction cleaved fragments may be blunt ended
5   by treating with the large fragment of E. coli DNA
polymerase I (Klenow reagent) in the presence of the four
deoxynucleotide triphosphates (dNTPs) using incubation
times of about 15 to 25 min at 20° to 25°C in 50 mM Tris
pH 7.6, 50 mM NaCl, 6 mM MgCl$_2$, 6 mM DTT and 0.1-1.0 mM
10 dNTPs. The Klenow fragment fills in at 5' single-stranded
overhangs in the presence of the four nucleotides. If
desired, selective repair can be performed by supplying
only one of the, or selected, dNTPs within the limitations
dictated by the nature of the overhang. After treatment
15 with Klenow reagent, the mixture is extracted with phenol/
chloroform and ethanol precipitated. Treatment under ap-
propriate conditions with S1 nuclease results in
hydrolysis of any single-stranded portions of DNA. In
particular, the nicking of 5' hairpins formed on synthesis
20 of cDNA is achieved.

Ligations are performed in 15-50 ul volumes
under the following standard conditions and temperatures:
for example, 20 mM Tris-Cl pH 7.5, 10 mM MgCl$_2$, 10 mM DTT,
33 mg/ml BSA, 10 mM-50 mM NaCl, and either 40 mM ATP,
25 0.01-0.02 (Weiss) units T4 DNA ligase at 14°C (for "sticky
end" ligation) or 1 mM ATP, 0.3-0.6 (Weiss) units T4 DNA
ligase at 14°C (for "blunt end" ligation). Intermolecular
"sticky end" ligations are usually performed at 33-100 mg/
ml total DNA concentrations (5-100 nM total end concentra-
30 tion). Intermolecular blunt end ligations are performed
at 1 mM total ends concentration.

In vector construction employing "vector frag-
ments", the vector fragment is commonly treated with
bacterial alkaline phosphatase (BAP) or calf intestinal
35 alkaline phosphatase (CIP) in order to remove the 5'

-36-

phosphate and prevent self-ligation of the vector. Diges-
tions are conducted at pH 8 in approximately 10 mM Tris-
HCl, 1 mM EDTA using about 1 unit per mg of BAP at 60°C
for one hour or 1 unit or CIP per mg of vector at 37°C for
about one hour. In order to recover the nucleic acid
5 fragments, the preparation is extracted with phenol/
chloroform and ethanol precipitated. Alternatively,
religation can be prevented in vectors which have been
double digested by additional restriction enzyme digestion
and separation of the unwanted fragments.

10

                          Example 1
        Coincidence Cloning of Blunt/Sticky-End Heteroduplexes

15   This example describes the use of coincidence
cloning to identify common genomic sequences in LAZ342, a
human lymphoblastoid cell line, and the somatic cell
hybrid HHW661, a hamster-human hybrid cell line containing
only a single human chromosome: a translocation of human
20 chromosome region 4p onto hamster chromosome 5. The
HHW661 cell line was prepared according to published
methods (Wasmuth).
            Genomic DNA from the two cell lines was obtained
by conventional methods (Maniatis), and both DNAs were cut
25 to completion with MboI, which generates fragments pre-
dominantly less than 1 kb in length. With reference to
Figure 1, the HHW661 DNA fragments were further blunt-
ended with Klenow fragment in the presence of all four
nucleotides, so that the final HHW661 fragments are blunt-
30 ended homoduplexes (DNA-I fragments in the figure) and the
lymphoblastoid cell fragments are sticky-ended (DNA-II in
the figure).
            Both mixtures of DNA fragments were mixed in a
1:1 ratio, alkaline denatured at pH 13, and then
35 reannealed by a phenol emulsion reassociation technique

-37-

(F-PERT) (Kohne; Casna). Specifically, the denatured DNA
fragments were mixed, and phenol and formamide were added
to final volume concentrations of 27 and 8 percent,
respectively. A two-phase emulsion was formed by vigorous
shaking with a vibratory shaker run at 1/2 to 3/4 maximum
5   speed. Total reannealing time was about 20-24 hours at
22° C. The annealed DNA was recovered by phenol extrac-
tion and ethanol precipitation, according to known
methods. As seen from Figure 1, the reannealed fragments
include original blunt-ended homoduplexes from the DNA-I
10   fragments, DNA-II homoduplexes with opposite MboI sticky
ends, and heteroduplexes with opposite MboI (or BamHI) and
blunt ends, as indicated. Reassociation of repeat
sequences in the fragment mixtures would not be expected
to yield clonable ends, since the repeats are likely to
15   hybridize with imperfect copies of themselves.

pUC18 plasmids were treated with BamHI and SmaI
restriction endonucleases, to cut the plasmid in its
polylinker region, and the small linker fragment was
removed by PEG precipitation. The reassociated fragments
20   from above were mixed and ligated with the cut plasmids
under standard conditions. Since the MboI ends of the
heteroduplex are compatible with the BamHI and SmaI ends
of the cut plasmid, respectively, only the heteroduplex
fragments are expected to form successful recombinants.
25   The recombinant plasmids are used to transform JM103 host
cells, and successful transformants are selected by plat-
ing in the presence of isopropylthiogalactoside (IPTG) and
5-bromo-4-chloro-3-methyl-indolyl-beta-D-galactoside
(Xgal). Minipreps of the plasmid DNA, designated pUC/HD
30   in Figure 1, revealed detectable inserts in the 200-1,000
bp range in 60% of the clones. Screening 48 colonies with
total hamster DNA by a colony filter hybridization
technique under conditions that only permit hybridization
of repeated sequences (Maniatis, p. 316) showed no
35   positives, and only 1 colony was positive for total human

-38-

DNA under the same conditions, indicating that the method
of the invention selects against repeated sequences.

Five inserts were purified from low-melting
temperature agarose (Sea Plaque) and labeled using random
hexamer priming (Feinberg). The five labeled fragments
5 were used as probes on Southern blots containing hamster
DNA, HHW661 DNA, and human DNA. Three of the five probes
gave single-copy bands in the HHW661 and human DNA lanes,
and no signal in the hamster lanes, indicating that these
fragments did in fact arise from the human translocation
10 chromosome, as expected.

Example 2
Coincidence Cloning with Mixed Sticky-End Linkers

15 Genomic DNA from the lymphoblastoid and HHW661
cell lines above is cut to completion with MboI, as
described, yielding predominantly 200-1,000 bp fragments
with MboI sticky ends, as illustrated in Figure 2, where
again the HHW661 fragments are indicated DNA-I and the
20 lymphoblastoid-cell fragments, as DNA-II.

Synthetic linkers having an MboI sticky end and
an internal XhoI site (linker I in the figure) or an
internal NotI site (linker II) are prepared by
conventional oligonucleotide methods, as described above.
25 The XhoI linker is ligated to the DNA-I fragments, and the
fragments are cut to completion with XhoI endonuclease,
yielding DNA-I fragments with XhoI sticky ends, as
indicated. Similarly, the DNA-II fragments are ligated
with the NotI linker, and the resulting fragments are cut
30 to completion with NotI endonuclease, yielding DNA-II
fragments with NotI sticky ends as shown. Here it is
noted that fragments with relatively rare internal XhoI or
NotI sites will not form the desired heteroduplexes
(below) and thus will be lost.

35

-39-

Both mixtures of DNA fragments are mixed in a
1:1 ratio, alkaline denatured at pH 13, then reannealed by
the formamid-phenol emulsion reassociation technique (F-
PERT) of Example 1. The annealed DNA is recovered by
phenol extraction and ethanol precipitation, as above. As
5   seen from Figure 2, the reannealed fragments include the
original homoduplexes from the DNA-I and DNA-II fragments,
having opposite XhoI and NotI ends, respectively, repeat
sequences with different-length ends, and heteroduplexes
with opposite XhoI and NotI ends.

10   A Bluescript® plasmid containing NotI and XhoI
sites in the plasmid's polylinker region is cut with XhoI
and NotI endonucleases, and the small linker fragment is
removed by polyethyleneglycol (PEG) precipitation. The
reassociated homoduplex and heteroduplex fragments from
15   above are mixed and ligated with the cut plasmids under
standard conditions. As can be appreciated from above,
and from Figure 2, only the end-hybridizable
heteroduplexes, with their opposite NotI and XhoI sticky
ends are compatible with the cut ends, of the plasmid, and
20   therefore only these heteroduplexes are expected to form
successful recombinants. Confirmation of successful re-
combinant plasmids is on JM103 host cells, with plating in
the presence of isopropylthiogalactoside (IPTG) and Xgal,
as above. Minipreps of the plasmid DNA are used to test
25   plasmids with inserts in the 200-1,000 bp size range.

Non-repeat clones are further purified and
labeled as above, for screening genomic fragments from
hamster, human and HHW661 cells, to identify those clones
which are specific for both human and HHW661 genomic frag-
30   ments, as determined, for example, by probe binding to
Southern blots of the genomic fragments.

35

-40-

Example 3
Coincidence Cloning with Methylated Heteroduplexes

Genomic DNA from the lymphoblastoid and HHW661
cell lines above is cut to completion with BamHI, yielding
5   fragments predominantly in the 2-10 kilobase size region
and having BamHI (B) sticky ends, as seen in Figure 3,
where the fragments derived from the HHW661 and
lymphoblastoid cell lines are designated DNA-I and DNA-II,
respectively. The DNA-I fragments are now treated with
10   AluI methylase, to block internal AluI (A) sites in the
fragments, and the DNA-II fragments are similarly treated
with HaeIII methylase, to block internal HaeIII (H) sites.
Restriction site methylation is indicated by the "*"
symbol on both fragment strands in the figure.

15   Both mixtures of DNA fragments are mixed in a
1:1 ratio, alkaline denatured at pH 13, then reannealed by
a phenol emulsion reassociation technique (F-PERT), as
above, and the annealed DNA is recovered by phenol extrac-
tion and ethanol precipitation, as above. As seen from
20   Figure 3, the reannealed fragments include the original
homoduplexes from the DNA-I and DNA-II fragments, having
opposite BamHI ends, same-size heteroduplex fragments also
having opposite BamHI ends, and unequal-strand homoduplex
and heteroduplex fragments (predominantly different-size
25   homologous repeat sequences) having at least one irregular
end.

The reannealed fragments are now digested to
completion with both AluI and HaeIII under standard digest
conditions to cut those fragments at internal, non-
30   methylated EcoRI and MboI sites, respectively. As can be
appreciated from Figure 3, AluI and HaeIII digestion of
the DNA-I homoduplexes, which have protected AluI sites,
produces BamHI/HaeIII (B/H) and HaeIII/HaeIII (H/H) frag-
ments in all fragments which have internal HaeIII sites.
35   Similarly, AluI and HaeIII digestion of the DNA-II

-41-

homoduplexes, which have protected HaeIII sites, produces
BamHI/AluI (B/A) and AluI/AluI (A/A) fragments in all
fragments which have internal AluI sites. Since the
equal-strand heteroduplex fragments are protected at both
sites, by methylation of the AluI sites on one strand and
5   the HaeIII sites on the homologous strand, the
heteroduplex is not susceptible to digestion and therefore
retains its two opposite BamHI sites. It is noted here
that the small percentage of homoduplex fragments which do
not contain internal AluI or HaeIII sites will also retain
10   their opposite BamHI sticky ends.
          The digest fragments from above are now ligated
into a pUC18 plasmid which has been linearized by BamHI
digestion. As above, the digested fragments are mixed and
ligated with the cut plasmids under standard conditions,
15   and the plasmids are selected for successful recombinants,
which should contain only the matched heteroduplex frag-
ments. Non-repeat clones are further purified and labeled
as above, for screening genomic fragments from hamster,
human and HHW661 cells, to identify those clones which are
20   specific for both human and HHW661 genomic fragments, as
determined, for example, by probe binding to southern
blots of the genomic fragments.

25   Example 4
Coincidence Cloning with Methylated-Linker Heteroduplexes

          Genomic DNA from the lymphoblastoid and HHW661
cell lines above is cut to completion with MboI or BamHI,
30   as above yielding predominantly 200-1,000 bp fragments
with MboI sticky ends, or 200-20,000 bp fragments with
BamHI sticky ends as illustrated in Figure 4. As in
Figures 1-3, the HHW661 fragments are indicated as DNA-I
and the lymphoblastoid-cell fragments, as DNA-II.

35

-42-

Synthetic linkers having an MboI (M) sticky end and internal HaeIII (H) and AluI (A) sites and either an XhoI (X) or a NotI (N) site adjacent the opposite linker end, are prepared by conventional methods, as detailed in the Materials and Methods section above. The nucleotide
5   sequence of the two linkers is shown in Figure 4. The XhoI linkers (linker I) are ligated to the DNA-I fragments, yielding fragments having groups of H/A/X sites at each end region. For purposes of illustration, the fragments illustrated in the figure are also shown as having
10   internal AluI (A) and HaeIII (H) sites, since in fact, many of the genomic fragments will contain such sites. The DNA-I fragments with attached linkers are now treated with AluI methylase, to methylate all AluI sites in the fragments, including those in the fragment end linkers.
15   As indicated by the "*" methylation symbol, both strands of the fragments are so methylated. Enzymatic conditions for ligating linkers to the DNA fragments and for methylating the fragments are conventional.

Similarly, the NotI (linker II) are ligated to
20   the DNA-II fragments, yielding fragments having groups of H/A/N restriction sites at each fragment end. Methylation of these fragments with HaeIII methylase gives the fragments indicated with methylated HaeIII sites in both of the DNA-II sites.

25   The mixtures of DNA fragments are mixed in a 1:1 ratio, alkaline denatured at Ph 13, then reannealed by a phenol emulsion reassociation technique (F-PERT), as above, and the annealed DNA is recovered by phenol extraction and ethanol precipitation, as above. As seen from
30   Figure 4, the reannealed fragments include the original homoduplexes from the DNA-I and DNA-II fragments, having either H/A/X or H/A/N linkers, respectively, at their opposite ends, repeat sequences with different-length ends, and heteroduplexes with an H/A/X linker sequence at one
35   end and an H/A/N linker sequence at the other end.

-43-

The reannealed fragments are now digested to
completion with both AluI and HaeIII endonucleases, under
standard digest conditions. With continued reference to
Figure 4, digestion of the AluI methylated homoduplexes
(the DNA-I homoduplexes) with the combination of
5   endonucleases cleaves the fragments at all HaeIII sites,
including the end linker sites, producing fragments whose
opposite ends have HaeIII blunt ends. Similarly, diges-
tion of the HaeIII methylated homoduplexes (the DNA-II
homoduplexes) with the combination of endonucleases
10 cleaves the fragments at all AluI sites, including the end
linker sites, producing fragments whose opposite ends have
AluI blunt ends. In the single-copy heteroduplex frag-
ments (formed from same-length strands), all of the AluI
and HaeIII sites are methylated on one strand or the
15 other, and so no endonuclease digestion occurs, yielding
intact heteroduplex fragments with opposite XhoI and NotI
ends. Duplex fragments which are not end-hybridized
sequences will be cleaved by the AluI or HaeIII
endonucleases only in duplexes where the homologous
20 strands are derived from the same original DNA mixture,
thus yielding fragments with irregular ends, or fragments
where one or both ends are AluI or HaeIII ends.

The digest fragments from above are now ligated
into a Bluescript® vector having NotI and XhoI polylinker
25 sites. Briefly, the vector is digested with the both NotI
and XhoI, with removal of the small linker fragment. As
above, the digested fragments are mixed and ligated with
the cut plasmids under standard conditions, and the
plasmids are selected for successful recombinants, which
30 should contain only the matched heteroduplex fragments.
Non-repeat clones are further purified and labeled as
above, for screening genomic fragments from hamster, human
and HHW661 cells, to identify those clones which are
specific for both human and HHW661 genomic fragments, as

35

-44-

determined for example, by probe binding to Southern blots
of the genomic fragments.

Example 5
Heteroduplex Selection of Mixed-Strand
Biotinylated Fragments: Method 1

Genomic DNA from the lymphoblastoid cell line
above is cut to completion with HindIII and EcoRI,
substantially as described, yielding predominantly 200-
10,000 bp fragments with HindIII and EcoRI sticky ends.
These fragments are then hybridized with single-copy,
biotinylated HHW661 DNA fragments also produced by
complete HindIII and EcoRI digestion, and prepared as
described in Parts A and B below.

A.   Removing Repetitive Sequences from the HHW661 DNA
Fragments

The HindIII/EcoRI digest fragments from the
HHW661 cell line are dissolved in 0.12 M phosphate buffer
containing 0.2 mM EDTA (PB). Repetitive-sequence DNA is
removed by standard hybridization methods which are
detailed in the literature (Britten). Briefly, the DNA is
raised to about $10^{\circ}C$ above the melting temperature (T_m),
as determined for example by absorption at OD_{260}. In the
buffer used above, the T_m is between about $80^{\circ}-90^{\circ}C$. The
material is then cooled slowly to about $25^{\circ}C$ below the T_m,
and allowed to anneal to a $C_o t$ value (mole/liter x sec) of
about 100, at which the repeat-sequence material is pre-
dominantly in reannealed form, and the non-repetitive
fraction, in denatured form. This duplex material is
separated from single-strand DNA by hydroxyapatite (HAP)
chromatography, according to standard procedures
(Britten). Briefly, HAP is suspended in 0.15 PB, 2 mM
EDTA, and poured into a water-jacketed column maintained
at the reannealing temperature. After washing the column

-45-

with several volumes of the reannealing buffer, the DNA
material is loaded onto the column and the single-strand
material eluted with several volumes of the buffer. This
material is combined, and precipitated with cold ethanol,
as above.

5   The precipitated single-strand material is
redissolved in annealing buffer, and the entire separation
procedure repeated, except that the reannealing is
performed at a temperature about 10° below the above T_m
value.

10 B.   Biotinylating the Single-Copy HHW661 Fragments
        The biotinylated nucleotides used are Bio-11-
dUTP (Brigati) which has an 11-atom linker arm separating
the biotin and the pyrimidine base, and Bio-19-SS-dUTP
(Herman) which has a 19-atom linker containing a disulfide
15 bond. ^{32}P-labeled dNTPs are included when monitoring of
the various steps of the method is desired. The labeled
nucleotides are incorporated into the double strand frag-
ments by one of the following methods:

20 1.   Nick-Translation
        A typical reaction, carried out in 60ml final
volume, contains 1 ug DNA in 50 mM Tris-Cl pH 7.5, 10mM
MgSO4, 0.1 mM DTT, 100 mM of each of the following
nucleotides: dATP, dGTP, and Bio-11-dUTP or Bio-19-SS-
25 dUTP, 5 uCi of [alpha-^{32}P] dCTP (Amersham, specific activ-
ity 3,000 Ci/mmole), 30 U DNA polymerase I, and 27 pg/ml
DNAse I. The reaction mixture is incubated at 14°C for
one hour, stopped by addition of EDTA to 10 mM and heated
at 68°C for 5 min. Labeled DNA is recovered by
30 chromatography over Sephadex G50 equilibrated and eluted
with 10 mM Tris-Cl, pH 7.5/1 mM EDTA (T.E.). When large
amounts of DNA are required, two to three nick-
translations are run in parallel and loaded onto one
column to obtain a concentrated DNA solution.
35

-46-

2.   Tailing by Terminal Transferase

This procedure is used only after the DNA is
first treated to produce 3' protruding ends (Maniatis).
The reaction mixture consists of 1 ug DNA in 100 mM potas-
sium cacodylate (pH 7.2), 2 mM $CoCl_2$, 0.2 mM DTT, 100 mM
5   Bio-11-dUTP, 50 mCi [alpha-^{32}P] dCTP, and 20 U terminal
transferase, added last. After incubation at 37°C for 45
min, an additional 20 U of enzyme is added and the incuba-
tion repeated. The reaction is terminated by EDTA added
to 10 mM, the DNA is recovered as described above,
10   precipitated with ethanol, washed with 70% ethanol and
resuspended in 50 ul buffer.

3.   Labeling by T4 DNA Polymerase Replacement Reaction

The reaction contains 1 ug of DNA in 33 mM Tris-
15   OAc (pH 7.9), 66 mM NaOAc, 10 mM MgOAc, 0.5 mM DTT, 0.1
mg/ml BSA, and 0.5 U T4 DNA polymerase. After incubation
at 37°C for 7 minutes, dATP, dGTP, and Bio-11-dUTP are
added to a final concentration of 150 mM, dCTP is added to
10 mM, 50 mCi of [alpha-^{32}P] dCTP (3000 Ci/mmole), and
20   TrisOAc, NaOAc, MgOAc, BSA, and DTT are added to maintain
previous concentrations. This reaction is incubated at
37°C for 30 min, then dCTP is added to a concentration of
150 mM, and the reaction incubated for an extra 60 min at
37°C. The reaction is stopped by addition of EDTA to 10
25   mM, heated at 65°C for 10 min, chromatographed and
processed as described before.

4.   Labeling by Photobiotinylation

This is carried out by standard procedures, as
30   outlined in the protocol supplied by the manufacturer
(Clontech, Palo Alto, CA).

C.   Selection of Mixed-Strand Heteroduplex Fragments

With continued reference to Figure 5, the
35   HindIII/EcoRI fragments from the human lymphoblastoid line

-47-

are cloned into the HindIII and EcoRI sites of M13, and
the plasmid is grown under conditions which produce M13
supernatant phage containing the inserts in single-strand
form. The phage material is harvested and mixed with the
biotinylated single-copy HHW661 fragments prepared as in
Parts A and B. The fragments are denatured and reannealed
using the F-PERT method referenced above.

        The reannealed material is passed over an avidin
column, for binding of biotinylated DNA to the column
material. A 1 ml silanized syringe plugged with silanized
glass wool is packed with 0.3 ml streptavidin-agarose and
washed with 0.15 PB, 2 mM EDTA. The hybridization mixture
from above is loaded onto the column which is then washed
with several volumes of the same buffer, to remove non-
hybridized cDNA.

        The material bound to the column is alkaline
denatured at pH 13, and the released (non-biotinylated)
DNA strands are eluted with the same high pH medium. The
non-biotinylated strand material which elutes is carried
in the single-stranded phage. This material, which
constitutes human lymphoblastoid DNA sequences which are
homologous to single-copy sequences from the HHW661
sequences, is transfected into the JM103 host, and grown
in either single-strand or double-strand form.

                        Example 6
        Heteroduplex Selection of Mixed-Strand
             Biotinylated Fragments: Method 2

        Genomic DNA from the lymphoblastoid cell line
and HHW661 line are each digested to completion with MboI,
and the HHW661 MboI fragments are biotinylated according
to Example 5A. The biotinylation is preferably carried
out by a method, such as nick translation or terminal
tailing with T4 DNA polymerase, which does not alter the
sticky end sequences of the fragments. Rather than

-48-

initially removing repeat-sequences from the mixture of the two fragments, the fragments are hybridized under F-PERT conditions, as above, yielding homoduplexes and heteroduplexes which contain both end-hybridized and non-end-hybridized fragments, as illustrated in Figure 6.

The reannealed material is fractionated on a streptavidin column as above, and non-biotinylated bound DNA strands are released by alkaline denaturation as above. The released single-strand species contain both single-copy and repeat-sequence strands which are in common between the two fragment mixtures.

The single-strand eluate fraction from above is ethanol precipitated and reannealed using the F-PERT procedure, resulting in two populations of double-strand fragments, as seen in Figure 5. These include fragments formed by reannealing of end-hybridizable strands, giving duplex fragments in which the original MboI ends are restored, and fragments formed by reannealing of non-end-hybridizable strands, which do not have defined sticky ends.

The reannealed fragments from above are mixed with pUC18 plasmid which has been linearized by digesting with BamHI, and the MboI-ended fragments are ligated into the cut plasmid according to standard procedures. Successful recombinants are selected as in Example 1, and the colonies are screened with repeat-sequence probes, also as above, to identify single-copy clones.

Example 7
Selection of Heteroduplexes with Mixed-Density Strands

Genomic DNA from the HHW661 cells is cut to completion with MboI, and the digest fragments are density labeled with N^{15} nucleotides which are incorporated into the two fragment strands by nick translation, or terminal tailing with T4 polymerase, according to methods described

-49-

above for biotinylating fragments. DNA from the
lymphoblastoid cell line above is cut to completion with
MboI for hybridization with the labeled HHW661 fragments.

The density-labeled MboI fragments (wavy-line
duplex fragments in Figure 7) are mixed with the unlabeled
lymphoblastoid-cell fragments (straight-line duplex frag-
ments in the figure), denatured, and reannealed using the
F-PERT method, as described above. As shown in the
figure, the annealing/reannealing process yields
homoduplexes labeled at neither or both strands, and
heteroduplexes labeled in one strand only. Among both
classes of fragments, homoduplex and heteroduplex, there
are matched-strand fragments formed predominantly from
single-copy, same-size strands, and unmatched-strand frag-
ments formed predominantly from repeat sequences with dif-
ferent sizes. As above, the matched-strand fragments will
have MboI sticky ends, whereas the unmatched-strand frag-
ments will not.

The reannealed mixture is fractionated by
equilibrium centrifugation in a CsCl gradient, according
to classical techniques (Meselson). At equilibrium, the
fragments will have partitioned into three gradient bands,
as indicated at the right in Figure 7. These three bands,
progressing toward greater density, are: unlabeled
homoduplexes; heteroduplexes (containing a single labeled
strand), and labeled homoduplexes. These bands are
identified by UV absorption, and the heteroduplex band is
removed by aspiration.

End-hybridized heteroduplex fragments are
selected by cloning into a vector BamHI site, as in
Example 5, and the cloned inserts may be further screened
to remove repeat sequences.

-50-

Example 8

Heteroduplex selection

Genomic DNA from the HHW661 and human
lymphoblastoid cell lines are digested to completion with
EcoRI and HindIII. The digest fragments from the HHW661

5   line (DNA-I fragments in Figure 8) are cloned into the
EcoRI/HindIII site of vector M13/mp19 which carries an
EcoI to HindIII orientation in its polylinker, to place
the fragments in a "5'-3'" orientation in the double-
strand vector. Similarly, the digest fragments from the

10 lymphoblastoid line (DNA-II fragments in Figure 8) are
cloned into the HindIII/EcoRI site of vector M13/mp19
which carries a HindIII to EcoRI orientation in its
polylinker, to place the fragments in a "3'-5'" orienta-
tion in the vector.

15   The two vectors with their two inserts are grown
under conditions of phage production, and the phage
harvested from the colony supernatant by conventional
methods. The phage from the M13/mp19 vector, which
produce the "plus" strand of the fragment are mixed with

20 the phage of the M13/mp18 vector, which produce the
"minus" strand of the fragment insert. The two phage
populations are rapidly annealed using the F-PERT method
described above. The duplex material, representing
homologous strands from the different DNA mixtures, is

25 separated from single-strand DNA by hydroxyapatite (HAP)
chromatography, according to standard procedures
(Britten). Briefly, HAP is suspended in 0.15 PB, 2 mM
EDTA, and poured into a water-jacketed column maintained
at the reannealing temperature. After washing the column

30 with several volumes of the reannealing buffer, the DNA
material is loaded onto the column and the single-strand
material is eluted with several volumes of the buffer.
The duplex material is eluted at elevated temperature with
buffer.

35

-51-

As seen in Figure 8, the heterologous duplex
fragments include end-hybridized inserts in which the vec-
tor polylinker sites EcoRI (R) and HindIII (H) on the op-
posite sides of the insert are aligned and homologous, and
non-end-hybridized fragments in which at least one of the
5   polylinker ends is unmatched. The heterologous duplex
material is now digested to completion with EcoRI and
HindIII to release end-hybridized heteroduplex inserts
with opposite EcoRI and HindIII ends, and these fragments
are cloned in the EcoRI/HindIII site of a pUC18 vector, as
10   above.   Alternatively, the relatively small EcoRI/
HindIII fragments can be separated by gel electrophoresis,
or by hybridization of the fragments with opposite-strand,
biotinylated M13 vector, and removal by streptavidin af-
finity chromatography, as in Example 5.

15

Example 9
Isolation of Unique Genomic Restriction Fragment
This section describes the isolation and cloning
of sequences from a unique SalI fragment from the human
genome. The method involves first performing a partial
20   digestion of the genome with SalI. The partial digest
fragments, which have size ranges from a few to up to
several thousand kilobases, are fractionated by pulsed-
field gel electrophoresis, and the gel is probed with a
radiolabeled, selected-sequence probe. Two of the gel
25   regions which are positive for hybridization to the probe
are eluted, digested completely with MboI, and the co-
incidence cloning method of the invention is used to
identify and isolate sequences from the unique SalI frag-
ment in each fragment mixture which binds to the selected
30   probe. Details of the method are as follows:

A.   Partial SalI Digestion
Peripheral blood lymphocytes (PBL) are pelleted
by low speed centrifugation, and washed two times with 10
35

-52-

ml of phosphate-buffered saline (PBS). The cells are suspended to a final concentration of about 1×10^7 cells/ml and a portion of the suspension is mixed with an equal volume of 1% low-gelling temperature agarose. The agarose mixture is cooled to 45-50° C and immediately pipetted

5   into a mold that makes 100 ul blocks, each about 2mm x 5 mm x 10 mm. The blocks are solidified by contacting the mold with ice.

The cells are disrupted in the agarose blocks by incubating the blocks for 2 days at 50° C with gentle

10   shaking in ESP buffer (0.5 M EDTA, pH 9.0, 1% sodium dodecyl sulfate (SDS), and 1 mg/ml proteinase K). After incubation the samples are stored at 4° C in ESP.

Prior to restriction endonuclease digestion, the blocks are treated with PMSF to inactivate proteases in

15   the block. This is done by treating each block twice with 1 ml of 1mM PMSF in TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 7.4), with slow rotation at room temperature for 2 hours. This is followed by three 1 ml washes with Tris-HCl buffer alone, for two hours each.

20   Partial digestions are carried out in 1.5 ml microfuge tubes containing 100 ug/ml bovine serum albumin and SalI in 10 mM Tris-HCl buffer, pH 7.4, to a final volume of 250 ul. The agarose blocks are added to the tubes before the addition of SalI. The final concentra-

25   tion of SalI is either 2, 5, or 10 units/ug DNA in the block. For blocks prepared as above and containing about 10 ug DNA, SalI is added to the tubes to a final amount of 10, 50, or 100 units. The tubes are incubated at 37° C for increasing time periods ranging from 30 minutes to 12

30   hours. To terminate the digestion, the buffer in a tube is carefully aspirated, and replaced with 1 ml of ES buffer (ESP without proteinase K), and the block is incubated in this buffer for 1 hour at 4°C. The buffer is then removed, replaced with 250 ul of ESP, and incubated

35   an additional 2 hours at 50°. After aspirating the

-53-

buffer, the block may be placed directly- in an agar slab
(below) for pulsed field gel electrophoresis (PFGE), or
stored at 4° C until use.

Optimal partial digest conditions are determined
by running each of the blocks from above on PFGE, and
5   determining the optional incubation period and SalI
concentration which give the desired size distribution of
partial digest fragments. As seen in Figure 9, under
optimal conditions, the genomic fragments will contain
between zero to 3 or more internal SalI (S) sites. For
10   purposes of illustration, the SalI fragment of interest in
the figure is the S_3/S_4 fragment, which is also contained
in the S_2/S_4 and S_1/S_4 fragments shown in the figure. In
particular, the S_3/S_4 fragment is a relatively large
genomic fragment which contains (a) a single-copy gene
15   sequence which is homologous to the labeled probe, and (b)
a gene region of interest. In general, the fragment of
interest is too large to clone as a single piece, and the-
probe-sequence region may be separated from the gene
region of interest by typically more than about 50 and up
20   to 1,000 kilobases.

B.   Size Fractionation by PFGE
The SalI partial digest fragments from above are
fractionated by PFGE, substantially according to published
25   methods (Smith; Schwartz). Briefly, a gel suspension
containing 1.0% agarose, and TBE buffer (10 mM Tris/Borate
buffer, pH 7.4 containing 0.1 mM EDTA) is poured into a 20
cm^2 mold to a depth of about 12 mm. After gel hardening,
slots corresponding in size to the gel blocks are cut
30   along one edge of the gel, and the gel blocks, typically
one every 2 cm, are placed in the slots. The slab is
placed in a horizontal gel box containing electrodes on
all four sides, at an angle of 45° with respect to the
sides of the box, i.e., such that the diagonals of the box

35

-54-

are normal to the sides of the slab. The arrangement of
electrodes described generally in Carle is used.

Electrophoresis is carried out with continuous
circulation of TBE buffer, with cooling of the circulated
buffer at 15°C, at pulse times of about 60 seconds at 200
5   volts. The electrophoretic run is terminated when the
marker bands have migrated to near the bottom of the gel,
as indicated by ethidium staining. Typical
electrophoresis times are between about 24 and 36 hours.
The gel is cut in half, providing one gel for use in
10 Southern blotting, and a second gel for use in obtaining
intact duplex SalI fragments. These two gels are referred
to below as "probe" and "recovery" gels, respectively.

C. Identifying Gel Bands

15   The probe and recovery gels from above are
stained by incubation in 1 ug/ml ethidium bromide for 10
minutes with gentle agitation on a platform shaker. The
gels are exposed briefly to a weak 360nm UV source during
which time photographs of the gel are taken. The two gels
20 are matched for corresponding stained regions, i.e., each
stained band in the recovery gel is matched with a cor-
responding band in the probe gel.

The probe gel is protected from light during
subsequent manipulations prior to and during Southern
25 blotting (Smith). Exposure to 254 nm UV light is for one
minute. Denaturation of gel DNA material is carried out
for one hour in 0.5 NaOH, 0.5M NaCl, and neutralization is
carried out for one hour in 1.5M Tris-Cl pH 7.5 with
gentle agitation. The gel is blotted to nitrocellulose by
30 ascending transfer overnight with a conventional sodium
citrate buffer (Maniatis). The filter is baked for two
hours in vacuo at 80°C, and stored in a tight container.
Using the labeled probe of interest, a Southern blot of
the gel fragments is prepared, according to standard
35 methods (Maniatis). From the blot, two probe-binding gel

-55-

band regions, such as the regions identified as containing fragments S_1/S_4 and S_3/S_4 in Figure 9, are identified. From the positions of these two gel regions, the corresponding regions in the recovery gel are removed for recovery of the fragments in each region. The fragment material is eluted from the gels by electroelution according to standard procedures, and the eluted DNA fragments are ethanol precipitated.

D. Cloning Single-Copy Sequences from the Selected Digest Fragment

The two SalI fragment mixtures obtained from the two gel regions above are each digested to completion with MboI, and the resulting fragments in one of the mixtures is further treated with Klenow fragment in the presence of all four nucleotides, as in Example 1, to fill in the sticky MboI ends. The two fragments mixtures are then mixed, denatured at pH 13, and reannealed by the F-PERT method, as in Example 1, to generate end-hybridized heteroduplexes which have opposite blunt and MboI sticky ends. The hybridization fragments are then cloned into the BamHI/SmaI site of pUC18, as in Example 1, and successful recombinants are identified and screened, both to remove repeat-sequence clones, and to identify clones which hybridize to the labeled probe used above to identify SalI fragments of interest on the PFGE gel. The methods of Examples 2-8 could also be applied.

Example 10
Removing Repeated Sequences from Genomic DNA Fragments

Genomic DNA is obtained from human PBLs as in Example 1, and this material is digested to completion with MboI as in Example 1. The digest material is divided into two equal portions, and one portion is further treated with Klenow fragment in the presence of all four

-56-

nucleotides, as in Example 1, to fill in the sticky MboI
ends. The two fragment mixtures are then mixed, denatured
at pH 13, and reannealed by the F-PERT method, as in
Example 1, to generate end-hybridized heteroduplex
sequences from the two mixtures which have opposite blunt
5 and MboI sticky ends. The hybridization products are then
cloned into the BamHI/SmaI site of pUC18, as in Example 1.
Successful recombinants can further be screened with
repeat-sequence probes to remove remaining repeat-sequence
clones in the library.

10   While the invention has been described with
reference to particular methods of coincidence cloning,
and applications of the method to specific problems in
genetic mapping and engineering, it will be apparent to
those skilled in the art that various alternative methods
15 and further applications may be developed
within the scope of the invention.

20

25

30

35

-57-

IT IS CLAIMED:

1. A method of obtaining, from a first mixture
of DNA duplex fragments derived from a first source, those
fragments which are homologous to and end-hybridizable
with the duplex DNA fragments in a second mixture of DNA
fragments derived from another source, where each duplex
is defined as having paired strands, said method compris-
ing

preparing the fragments from at least one of the
mixtures so that when a fragment strand from the one
mixture is hybridized with a homologous, end-hybridizable
strand from the other mixture, the resulting end-
hybridized fragment has properties which allow its isola-
tion from homoduplex fragments produced by hybridization
between opposite strands of the fragments in the first or
second mixture only, and from heteroduplex fragments which
are not end-hybridized,

reacting opposite strands from the fragments of
the first and second mixtures in a reaction mix under
hybridization conditions which yield heteroduplex frag-
ments, and

isolating the end-hybridized heteroduplex frag-
ments from other nucleic acid species contained in the
reaction mix.

2. The method of claim 1, wherein said isolat-
ing includes introducing the fragments produced by said
reacting into a cloning vector which selectively in-
corporates those end-hybridized heteroduplex fragments.

3. The method of claim 2, wherein the fragments
in the two mixtures are generated under conditions which
yield a pair of ligatable ends in the end-hybridized
heteroduplex fragments which is different from the pair of

-58-

ligatable ends in either the first-mixture or second-mixture homoduplexes.

4. The method of claim 3, wherein the duplex fragments in the first mixture are prepared by (a) cutting duplex DNA with a selected restriction endonuclease which produces fragments with sticky ends, and (b) blunt ending the sticky fragment ends, the fragments in the second mixture are prepared by cutting duplex DNA with a selected restriction endonuclease which also produces fragments with sticky ends, and the heteroduplex fragments have opposite blunt and sticky ends, and the homoduplexes have either opposite blunt or opposite sticky ends.

5. The method of claim 2, wherein said preparing includes attaching to the fragments in the first and second fragment mixtures, end linkers which can be manipulated to yield one ligatable end A at the opposite ends of the first-mixture fragments, and a second ligatable end B at the opposite ends of the second-fragment mixture, and the end-hybridized heteroduplex fragments have A and B ligatable ends at their opposite fragment ends.

6. The method of claim 5, wherein the linkers attached to the first- and second-mixture fragments have internal A and B sequences, respectively, and said preparing further includes cutting each of the first- and second-mixture fragments with an endonuclease which is specific for the A and B sequence, respectively.

7. The method of claim 2, wherein said preparing includes attaching to the first-mixture fragments, an end linker having restriction site sequences A, B, and C, where A and B are internal to C when the linker is attached to the fragments; attaching to the second-mixture

-59-

fragments, an end linker having restriction site sequences
A, B, and D, where A and B are internal to D when the
linker is attached to the fragments, and C and D are dif-
ferent sequences; treating the first-mixture fragments
with a methylase specific for sequence A in the fragments,
5 and treating the second-mixture fragments with a methylase
specific for the B sequence in the fragments; and said
isolating includes digesting the duplex fragments produced
by said reacting with endonucleases which cut the frag-
ments at non-methylated A and B sequences, and cloning the
10 fragments into a cloning vector which incorporates
selectively fragments with opposite-end C and D ligatable
ends.

   8.   The method of claim 1, wherein said prepar-
15 ing includes incorporating into the two strands of the
fragments in one of the mixtures, a label which allows
physical separation of heteroduplex fragments containing
one labeled strand from those in which either both or
neither fragment strands contain such label.

20
   9.   The method of claim 8, wherein the label is
an epitopic molecule, and said isolating includes contact-
ing the homoduplex and heteroduplex fragments produced by
said reacting with an affinity support material containing
25 a binding molecule capable of binding specifically and
.with high affinity to the epitopic molecule, to bind frag-
ments in which either one or both strands contain the
epitopic label, and treating the support material and at-
tached fragments to denature the fragments and release
30 strands which do not contain the epitopic label, where the
pairs of epitopic molecule/binding molecule are selected
from the group consisting of biotin/avidin, biotin/
streptavidin, carbohydrate/lectin, and antigen/antibody
pairs.

35

-60-

10.   The method of claim 9, wherein one of the two fragment mixtures has been prepared to remove repeat-sequence fragments, and the fragment mixture which is not labeled is cloned into a single-strand cloning vector, wherein the unlabeled fragment strands released from the support material can be used to transfect a suitable host, for growth in either single-strand or double-strand form.

11.   The method of claim 9, wherein said isolating includes reannealing the unlabeled fragment strands released from the support material, to produce homologous fragments derived from the unlabeled fragment mixture and homologous to sequences in the first mixture, and introducing said reannealed homologous fragments into a cloning vector which selectively incorporates those duplex fragments having ligatable ends.

12.   The method of claim 8, wherein said label is an isotopically labeled nucleotide which increases the buoyant density of duplex fragments containing one or both labeled strands, and said isolating includes separating heteroduplex from homoduplex fragments by density centrifugation.

13.   The method of claim 12, wherein said isolating further includes introducing heteroduplex frag-ments produced by said reacting, and having ligatable ends, into a cloning vector.

14.   The method of claim 1, wherein said prepar-ing includes cloning the DNA fragments of the first and second mixtures into cloning vectors which can be grown under conditions which yield single strand vectors containing only the sense strands from the first-mixture fragments, and only the anti-sense strands from the second-mixture fragments, said reacting produces

-61-

heteroduplex fragments only, and said isolating includes
separating duplex from non-duplex DNA species.

15. The method of claim 14, wherein the first-
and second-mixture fragments are cloned in opposite
orientations into a single-strand cloning vector, yielding
vector sequences A and B on either side of the inserted
fragments, and said isolating further includes digesting
the heteroduplex fragments with endonucleases with cut at
or adjacent said A and B sequences, and cloning the
heteroduplex fragments released by said digesting into a
vector.

16. The method of claim 1, for use in cloning
one or more regions of a DNA restriction fragment
contained in a mixture of restriction fragments generated
from a DNA source and containing a region which is
homologous to a selected DNA probe, comprising
        generating the restriction fragments by
endonuclease digestion of the DNA source,
        fractionating the resulting DNA fragments into
several subfractions,
        identifying two different subfractions which
each bind to a selected probe which is homologous to a
region in the restriction fragment of interest, and
        applying the method of claim 1 to the two
subfractions, to obtain those DNA sequences in the first
subfraction which are homologous to sequences in the
second subfraction.

17. The method of claim 16, wherein said
generating includes partially digesting the source DNA
with a rare-cutter endonuclease, said fractionating is
performed by pulsed-field gel electrophoresis, said
obtaining includes identifying two gel band regions which
bind to the probe of interest, and eluting fragments from

-62-

these two regions, and said applying includes digesting
the fragments in the two subfractions with one or more
restriction endonucleases which reduce the average sizes
of the fragments in the mixture to less than about 20
kilobases.

5

18. The method of claim 1, for use in cloning
conserved gene sequences from two different species,
comprising
        isolating the genomic DNA from the two different
10 species,
        digesting the genomic DNA from the two species
with one or more selected endonucleases, to produce first
and second mixtures of genomic DNA fragments from the two
different species, and
15   applying the method of claim 1 to the two
mixtures of DNA fragments, to obtain those DNA sequences
in the first mixture which are homologous to sequences in
the second mixture.

20   19. The method of claim 1, for use in enriching
a DNA fragment mixture with single-copy sequences,
comprising
        dividing the mixture into two portions, and
        applying the method of claim 1 to the two por-
25 tions of DNA fragments.

        20. The method of claim 1, for use in obtaining
cloned sequences from a selected chromosome or chromosome
region, comprising
30   providing a first hybrid cell line which
contains such chromosome or chromosome region,
        providing a second hybrid cell line which also
contains such chromosome or chromosome region, where the
two cell lines do not have any other common-species
35 chromosomes,

-63-

obtaining the genomic DNA from the first and
second cell lines, and digesting the DNA with one or more
restriction endonucleases, to produce first and second
mixtures of DNA fragments, respectively, and

applying the method of claim 1 to the two

5   mixtures of DNA fragments, to obtain those DNA sequences
in the first mixture which are homologous to sequences in
the second mixture.

21.   A library of cloned DNA sequences produced

10   by treating two DNA fragment mixtures according to the
method of claim 2.

22.   The library of claim 21, wherein one of the
strands in each heteroduplex fragment contains a

15   nucleotide label not present in the other homologous
strand.

23.   The composition of claim 22, wherein the
label is selected from the group consisting of

20   biotinylated, density-labeled, and methylated nucleotides.

25

30

35
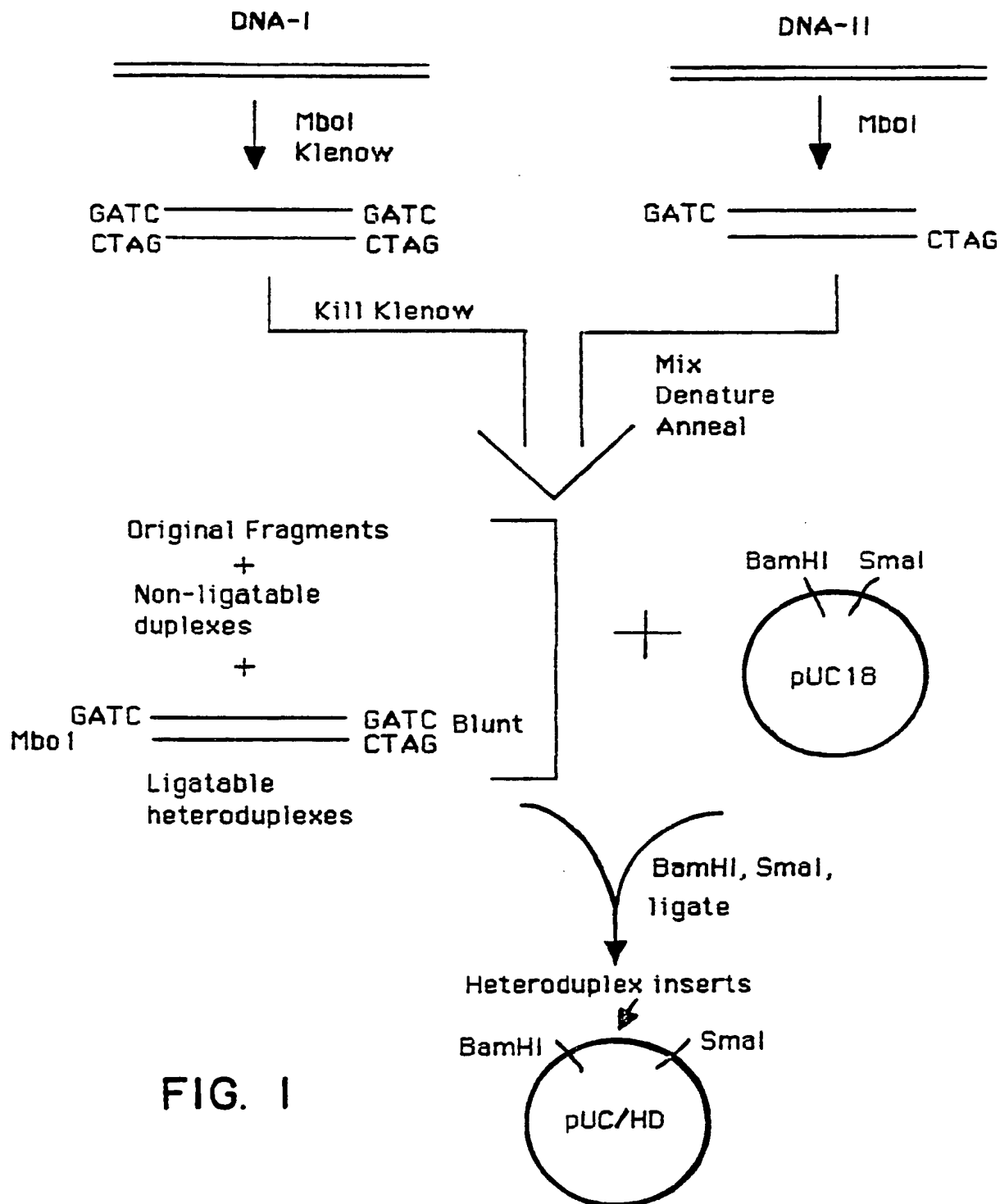
FIG. 1

FIG. 2

3/10

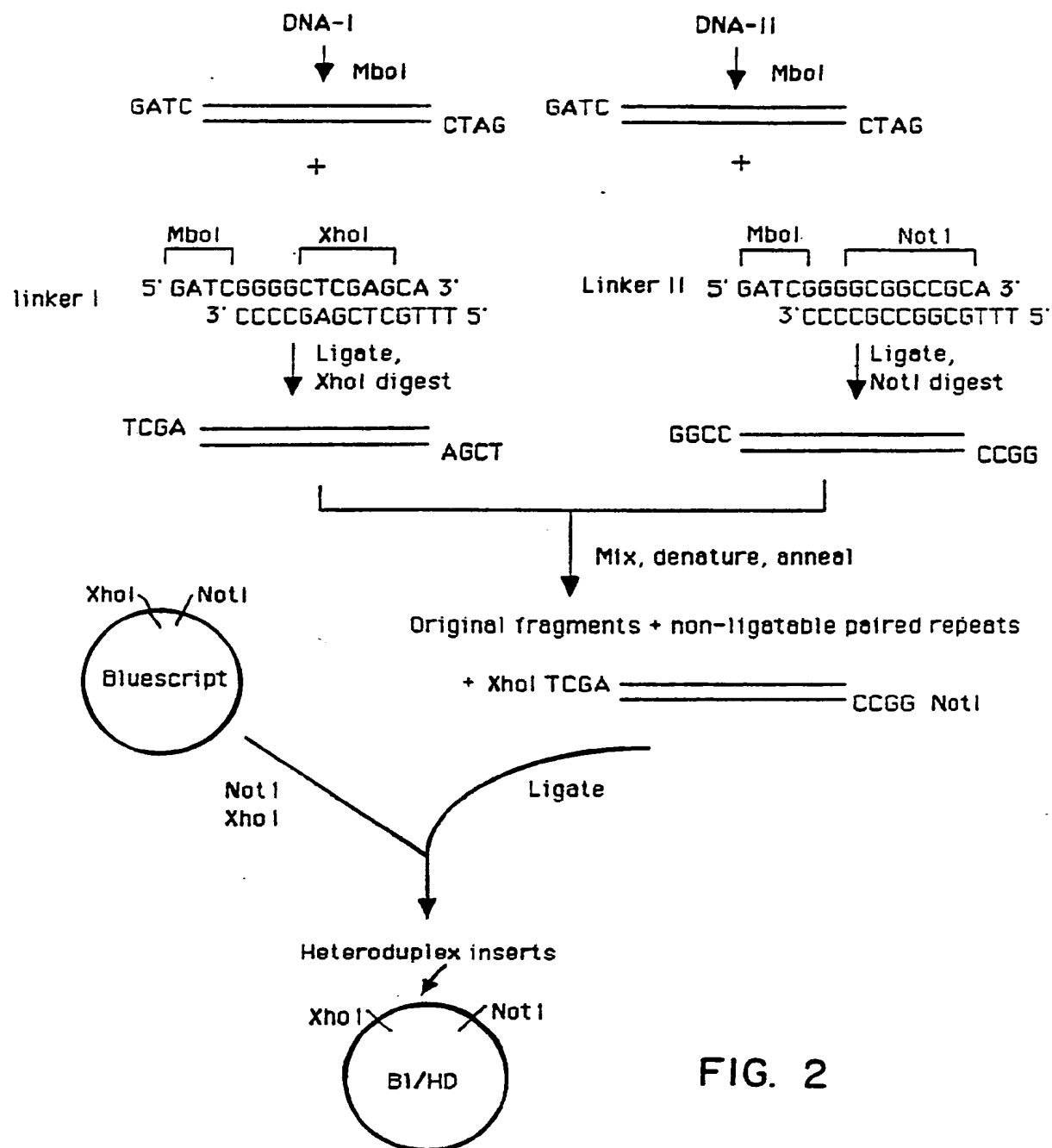FIG. 3

DNA-I   DNA-II

↓ Mbol   ↓ Mbol

GATC ———————— CTAG   GATC ———————— CTAG

A   H H   A   H H

+   +

HaeIII(H)   XhoI (X)

Mbol(M)   AluI(A)

linker I

GATCGGCCAGCTGCTCGAGCA
CCGGTCGACGAGCTCGTTT

Mbol HaeIII AluI   (N) NotI

linker II

GATCGGCCAGCTGCGGCCGCA
CCGGTCGACGCCGGCGTTT

↓ Ligate, Xhol digest   ↓ Ligate, Notl digest

AH   HA   AH   HA

X ———————— X   N ———————— N

A   H   H   A   H   H

↓ AluI Methylase   ↓ HaeIII Methylase

AH*   *   HA*   AH*   *   *   HA*

X ———————— X   N ———————— N

*   A*   H   H   *   *   A   H*   H*   *

↓ Mix, denature, anneal

homoduplexes

+

*

————————   Non-ligatable
         heteroduplexes
 *

+

*AH   *   HA*

X ———————— N   Ligatable heteroduplexes

*   A   H*   H*   *

↓ Fig. 4 con't

FIG. 4

(*indicates methylation of
one or both strands at site
of AluI on HaeIII)
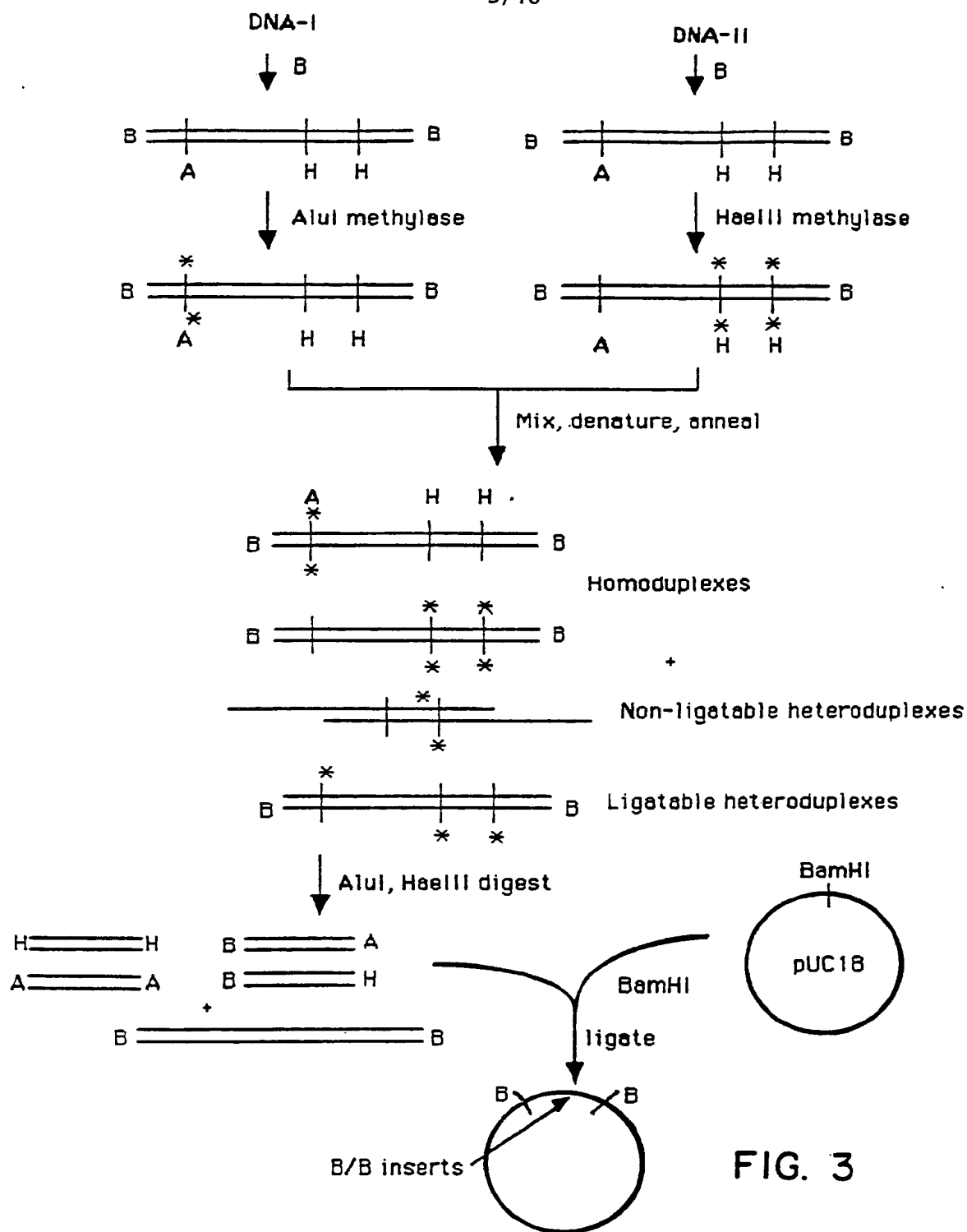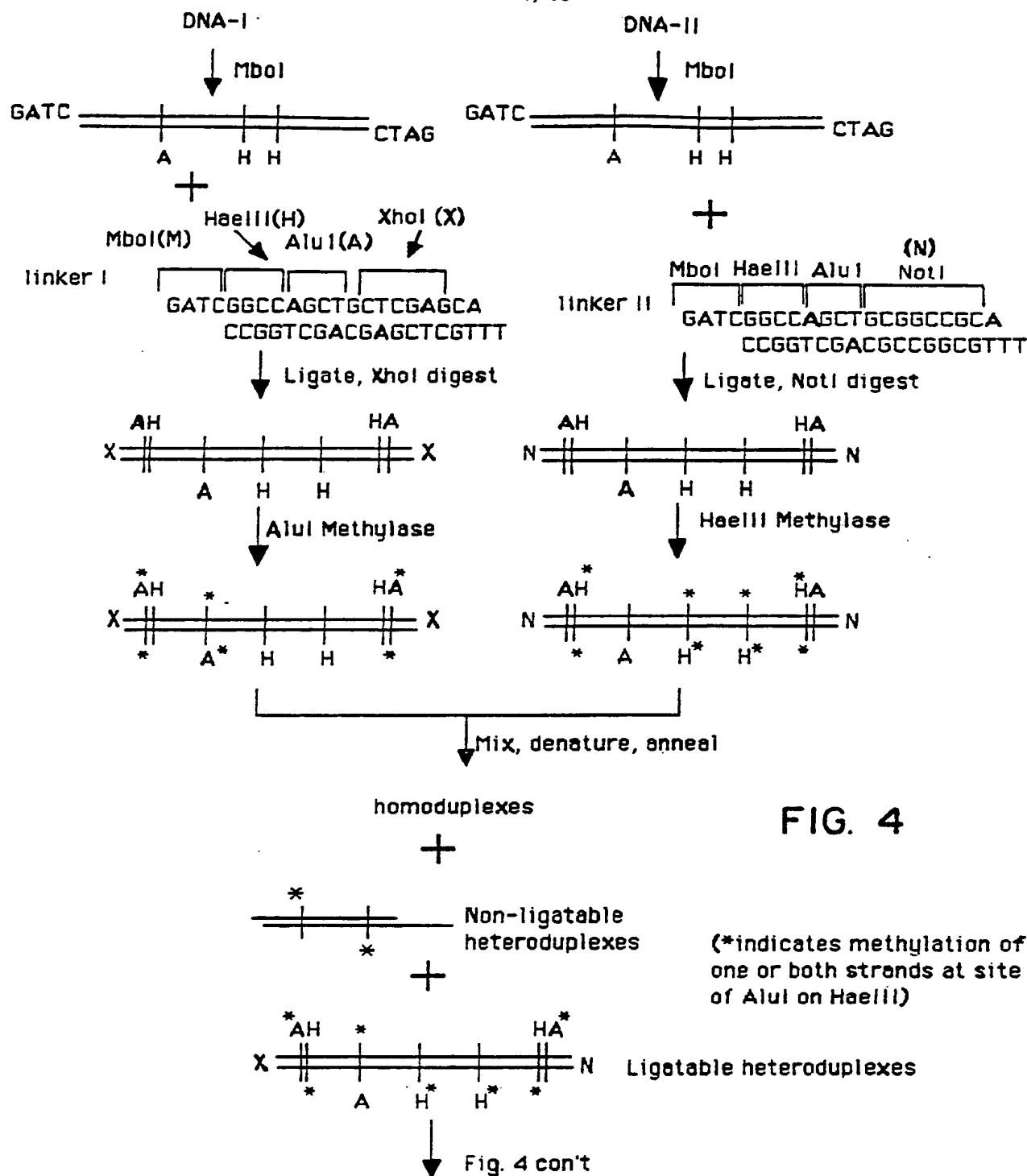
Fig. 4 Con't

Alul, HaeIII digestion

FIG. 4 (cont.)

A ══════ A

H ═════ H

X ═══════════ N

Xho!   Not!

Bluescript

Xhol, Notl,
remove polylinker

heteroduplex insert

X   N

FIG. 7

Denature,
anneal

Density gradient
centrifugation

FIG. 5

DNA (I)   DNA (II)

↓ Biotinylate

FIG. 6

Denature,
anneal

homoduplexes

+

heteroduplexes

Avidin column
Heat denature, elute

5'————————————————3'   heteroduplex strands
3'————————————————5'

Reanneal

Mbol ══════════════════ Mbol + ════════════════════

Ligatable fragments   Non-ligatable fragments

Clone into Mbol Vector site

SUBSTITUTE SHEET

FIG. 8

FIG. 9

FIG. 10

Cloned telomere sequences

INTERNATIONAL SEARCH REPORT

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 6

According to International Patent Classification (IPC) or to both National Classification and IPC

Int Cl. (4): C12Q 1/68
U.S. CL.: 435/6

II. FIELDS SEARCHED

Minimum Documentation Searched 7

Classification System	Classification Symbols
US.	435/6,803,172.3,253,320 436/501; 935/78 536/27

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched 8

Computer Searches: APS, Chemical Abstracts; BIOSIS, Derwent
(WPI, WPIL, Biotechnology)

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
P,Y	US, A, 4,710,465 (WEISSMAN ET AL.) 01 December 1987, see column 17, lines 18-36.	8-11,22 23
Y	Gene, Volume 24, issued September 1984 (Amsterdam, Netherlands), CHANG, S. ET AL. "Recombination following transformation of Escherichia coli by heteroduplex plasmid DNA molecules" see pages 255, 257, 259 and 260.	1-23
Y	Nucleic Acids Research, Volume 14, issued September 25, 1986 (Oxford, England) CASNA. N.J. ET AL "Genomic analysis II: isolation of high molecular weight heteroduplex DNA following differential methylase protection and Formamide-PERT hybridization" see pages 7285, 7299 and 7301.	1-23

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
30 October 1988	2·3 NOV 1988
International Searching Authority	Signature of Authorized Officer
ISA/US	JEREMY M. JAY

III. D CUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	US, A, 4,594,318 (GUSELLA ET AL) 10 June 1986, see abstract.	20
Y	DNA, Volume 3, issued February, 1984 (N.Y., USA) STUDENCKI, A ET AL "Allele-specific hybridization using oligonucleotide probes of very high specific activity: Discrimination of the human beta-A- and beta-S- globin genes" see page 10.	8,22
Y	Recombinant DNA a short course, issued 1983, WATSON, J.D. ET AL., W.H. Freeman and Company (N.Y. U.S.A) see page 51.	12,13,23